NTOU

National Taipei University

Artificial Intelligence for Text Analytics in FinTech (金融科技人工智慧文本分析)

Host: 林川傑 助理教授 (Prof. Chuan-Jie Lin)
Department of Computer Science and Engineering, National Taiwan Ocean University (NTOU)
Time: 13:00-15:00, May 19, 2022 (Thursday)
Place: Microsoft Teams, NTOU

戴敏育 副教授

Min-Yuh Day, Ph.D, Associate Professor

國立臺北大學 資訊管理研究所

Institute of Information Management, National Taipei University

https://web.ntpu.edu.tw/~myday

2022-05-19

戴敏育 博士
(Min-Yuh Day, Ph.D.)

國立臺北大學 資訊管理研究所 副教授
中央研究院 資訊科學研究所 訪問學人
國立臺灣大學 資訊管理 博士

Publications Co-Chairs, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013-)

Program Co-Chair, IEEE International Workshop on Empirical Methods for Recognizing Inference in TExt (IEEE EM-RITE 2012-)

Publications Chair, The IEEE International Conference on Information Reuse and Integration for Data Science (IEEE IRI)

國立臺北大學
National Taipei University

aws academy
Accredited Educator

aws certified
Solutions Architect
Associate

aws certified
Cloud Practitioner

aws educate
Cloud Ambassador
2020 Cohort

Outline

- AI for Text Analytics
  - Natural Language Processing with Transformers: Building Language Applications with Hugging Face
  - Practical Natural Language Processing
- FinTech: Financial Services Innovation
- Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech

AIWISFIN

AI Conversational Robo-Advisor (人工智慧對話式理財機器人)

First Place, InnoServe Awards 2018

https://www.youtube.com/watch?v=sEhmyoTXmGk

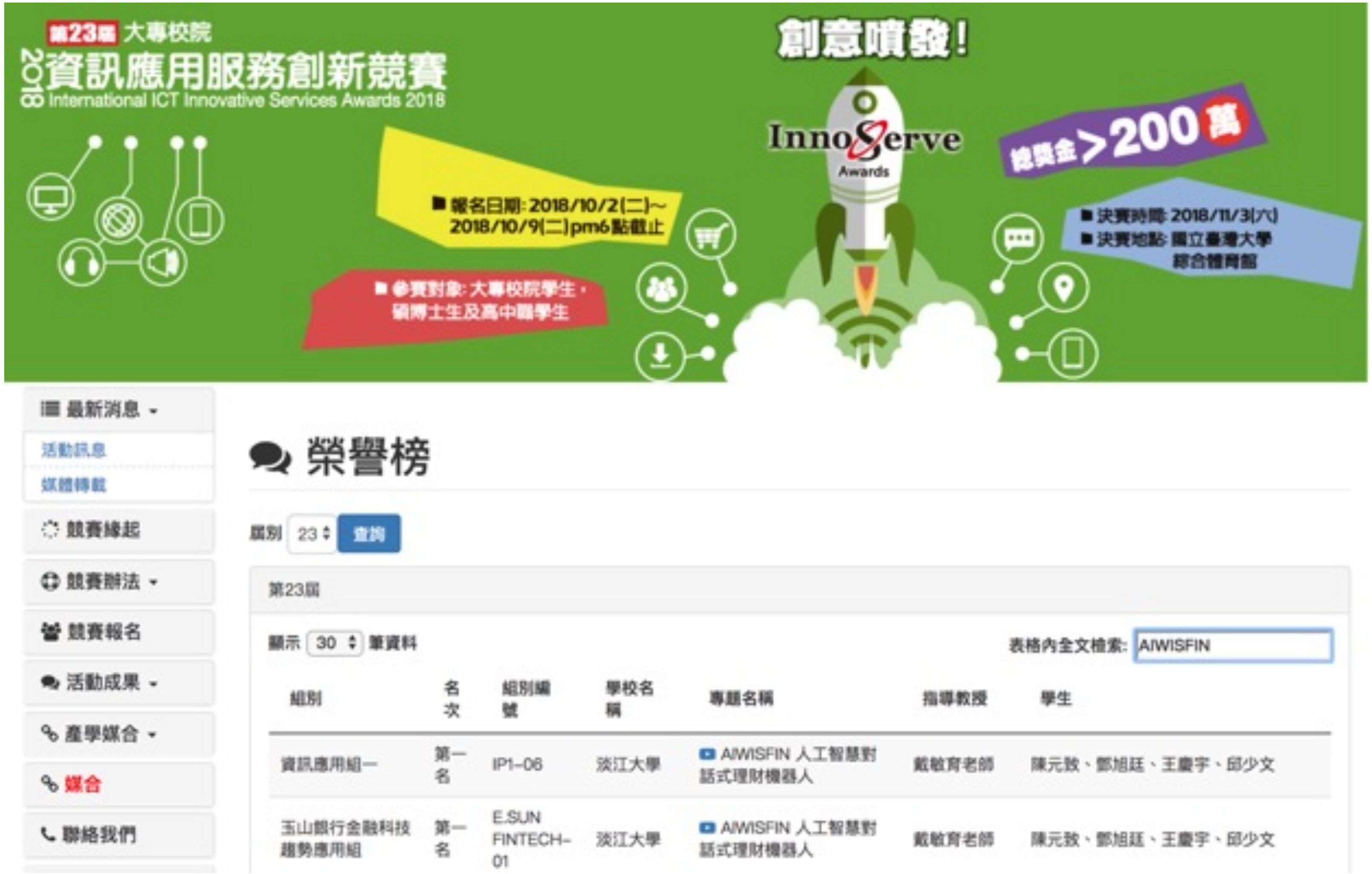
2018 The 23rd International ICT Innovative Services Awards (InnoServe Awards 2018)

InnoServe Awards

- Annual ICT application competition held for university and college students
- The largest and the most significant contest in Taiwan.
- More than ten thousand teachers and students from over one hundred universities and colleges have participated in the Contest.

https://innoserve.tca.org.tw/award.aspx

2018 International ICT Innovative Services Awards (InnoServe Awards 2018) (2018第23屆大專校院資訊應用服務創新競賽)



https://innoserve.tca.org.tw/award.aspx

Tamkang University

淡江大學

2011

Tamkang University

IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-9 RITE

Department of Information Management
Tamkang University, Taiwan

Min-Yuh Day

Chun Tu

myday@mail.tku.edu.tw

NTCIR-9 Workshop, December 6-9, 2011, Tokyo, Japan

Tamkang University
淡江大學

2013

Tamkang University

IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-10 RITE-2

Department of Information Management
Tamkang University, Taiwan


Min-Yuh Day


Chun Tu


Hou-Cheng Vong


Shih-Wei Wu


Shih-Jhen Huang

myday@mail.tku.edu.tw

NTCIR-10 Conference, June 18-21, 2013, Tokyo, Japan

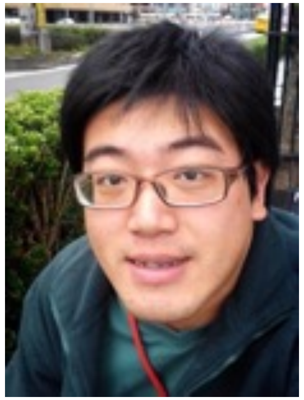
IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-11 RITE-VAL

2014

Tamkang University
淡江大學

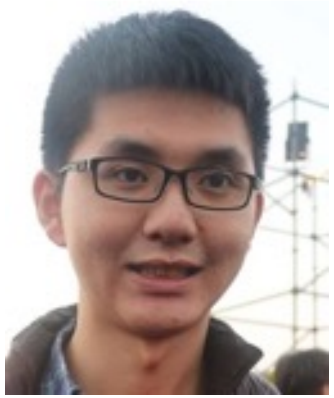

Min-Yuh Day


Ya-Jung Wang


Che-Wei Hsu


En-Chun Tu


Huai-Wen Hsu


Yu-An Lin


Shang-Yu Wu


Yu-Hsuan Tai


Cheng-Chia Tsai

NTCIR-11 Conference, December 8-12, 2014, Tokyo, Japan

NTCIR

Tamkang University
淡江大學

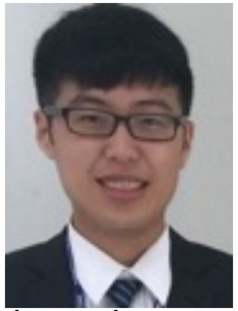
2016

IMTKU Question Answering System for World History Exams at NTCIR-12 QA Lab2

Department of Information Management
Tamkang University, Taiwan

Sagacity Technology


Min-Yuh Day


Cheng-Chia Tsai


Wei-Chun Chung


Hsiu-Yuan Chang


Tzu-Jui Sun


Yuan-Jie Tsai


Jin-Kun Lin


Cheng-Hung Lee


Yu-Ming Guo


Yue-Da Lin


Wei-Ming Chen


Yun-Da Tsai


Cheng-Jhih Han


Yi-Jing Lin


Yi-Heng Chiang


Ching-Yuan Chien

myday@mail.tku.edu.tw

NTCIR

2017

Tamkang University

淡江大學

IMTKU Question Answering System for World History Exams at NTCIR-13 QALab-3

Department of Information Management

Tamkang University, Taiwan

Min-Yuh Day

Chao-Yu Chen

Wanchu Huang

Shi-Ya Zheng

I-Hsuan Huang

Tz-Rung Chen

Min-Chun Kuo

Yue-Da Lin

Yi-Jing Lin

myday@mail.tku.edu.tw

NTCIR-13 Conference, December 5-8, 2017, Tokyo, Japan

NTCIR

2019

Tamkang University
淡江大學

IMTKU Emotional Dialogue System for Short Text Conversation at NTCIR-14 STC-3 (CECG) Task

Department of Information Management
Tamkang University, Taiwan


Min-Yuh Day


Chi-Sheng Hung


Yi-Jun Xie


Jhih-Yi Chen


Yu-Ling Kuo


Jian-Ting Lin

myday@mail.tku.edu.tw

NTCIR-14 Conference, June 10-13, 2019, Tokyo, Japan

NTCIR

2020

Tamkang University
淡江大學

IMTKU Multi-Turn Dialogue System Evaluation at the NTCIR-15 DialEval-1 Dialogue Quality and Nugget Detection

[1] Zeals Co., Ltd. Tokyo, Japan
[2] Information Management, Tamkang University, Taiwan
[3] Information Management, National Taipei University, Taiwan

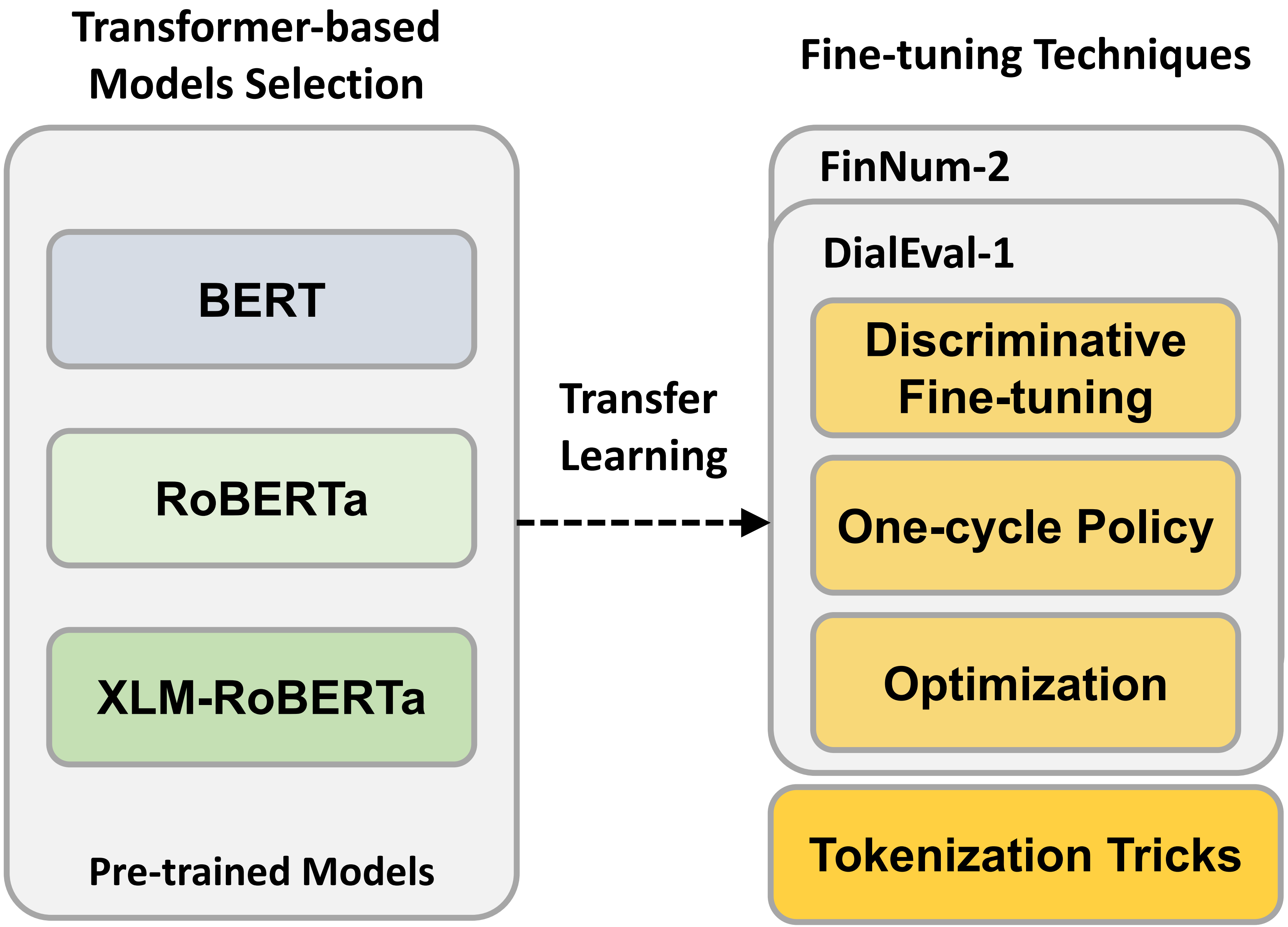
Mike Tian-Jian Jiang[1] Zhao-Xian Gu[2] Cheng-Jhe Chiang[2] Yueh-Chia Wu[2] Yu-Chen Huang[2] Cheng-Han Chiu[2] Sheng-Ru Shaw[2] Min-Yuh Day[3]

2020 NTCIR-15 Dialogue Evaluation (DialEval-1) Task
Dialogue Quality (DQ) and Nugget Detection (ND)

Chinese Dialogue Quality (S-score) Results (Zeng et al., 2020)

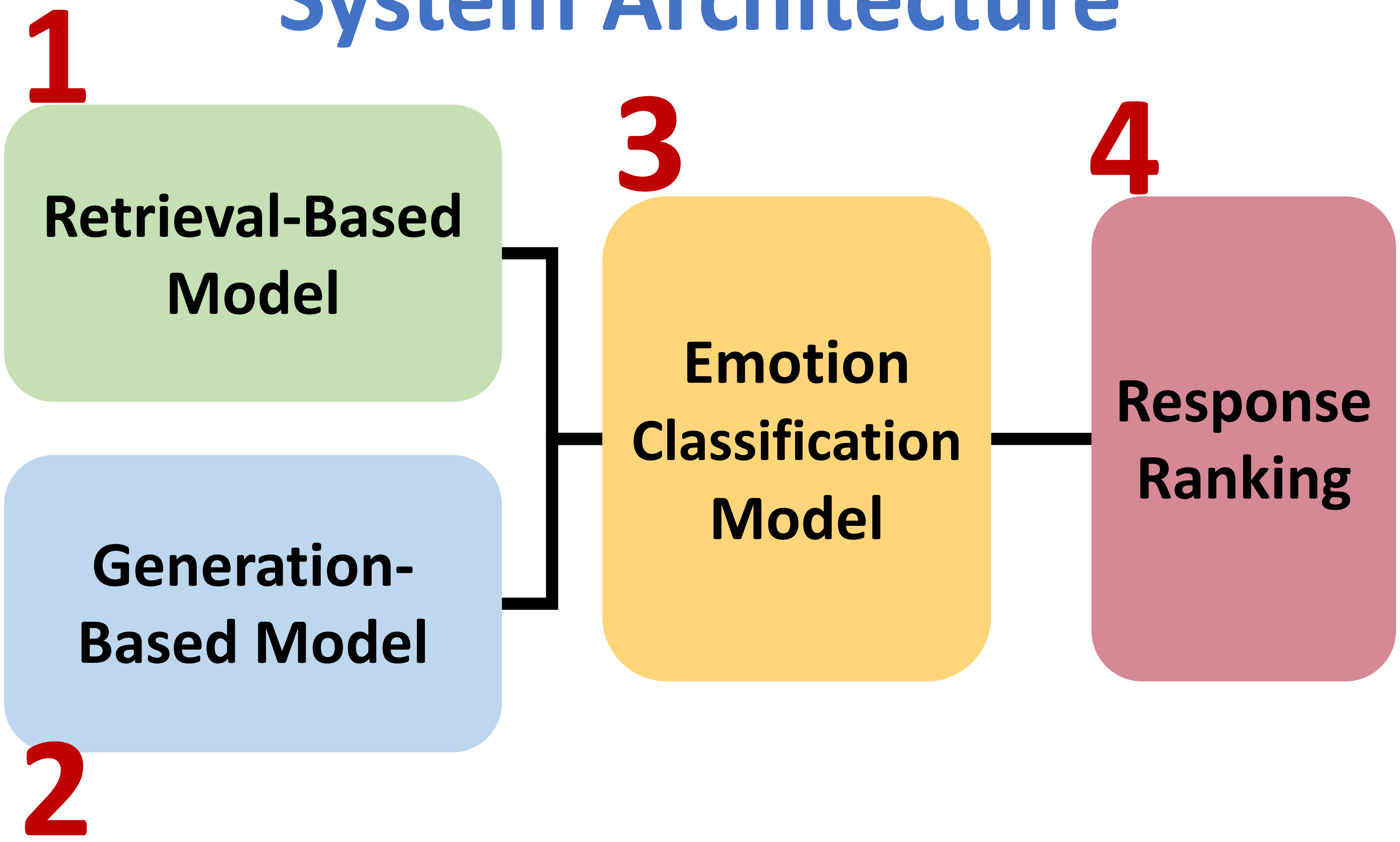
Run	Mean RSNOD	Run	Mean NMD
IMTKU-run2	0.1918	IMTKU-run2	0.1254
IMTKU-run1	0.1964	IMTKU-run0	0.1284
IMTKU-run0	0.1977	IMTKU-run1	0.1290
TUA1-run2	0.2024	TUA1-run2	0.1310
TUA1-run0	0.2053	TUA1-run0	0.1322
NKUST-run1	0.2057	NKUST-run1	0.1363
BL-lstm	0.2088	TUA1-run1	0.1397
WUST-run0	0.2131	BL-popularity	0.1442
RSLNV-run0	0.2141	BL-lstm	0.1455
BL-popularity	0.2288	RSLNV-run0	0.1483
TUA1-run1	0.2302	WUST-run0	0.1540
NKUST-run0	0.2653	NKUST-run0	0.2289
BL-uniform	0.2811	BL-uniform	0.2497

Source: Zeng, Zhaohao, Sosuke Kato, Tetsuya Sakai, and Inho Kang (2020), "Overview of the NTCIR-15 Dialogue Evaluation (DialEval-1) Task", Proceedings of NTCIR-15, 2020



Source: Jiang, Mike Tian-Jian, Shih-Hung Wu, Yi-Kun Chen, Zhao-Xian Gu, Cheng-Jhe Chiang, Yueh-Chia Wu, Yu-Chen Huang, Cheng-Han Chiu, Sheng-Ru Shaw, and Min-Yuh Day (2020). "Fine-tuning techniques and data augmentation on transformer-based models for conversational texts and noisy user-generated content." In 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 919-925. IEEE, 2020.

IMTKU Emotional Dialogue System Architecture



Short Text Conversation Task (STC-3) Chinese Emotional Conversation Generation (CECG) Subtask

Source: http://coai.cs.tsinghua.edu.cn/hml/challenge.html

NTCIR Short Text Conversation
STC-1, STC-2, STC-3

	Japanese	Chinese	English
NTCIR-12 STC-1 22 active participants	Twitter, Retrieval	Weibo, Retrieval	
NTCIR-13 STC-2 27 active participants	Yahoo! News, Retrieval+ Generation	Weibo, Retrieval+ Generation	
NTCIR-14 STC-3		Weibo, Generation for given emotion categories	
		Weibo+English translations, distribution estimation for subjective annotations	

Chinese Emotional Conversation Generation (CECG) subtask

Dialogue Quality (DQ) and Nugget Detection (ND) subtasks

Single-turn, Non task-oriented

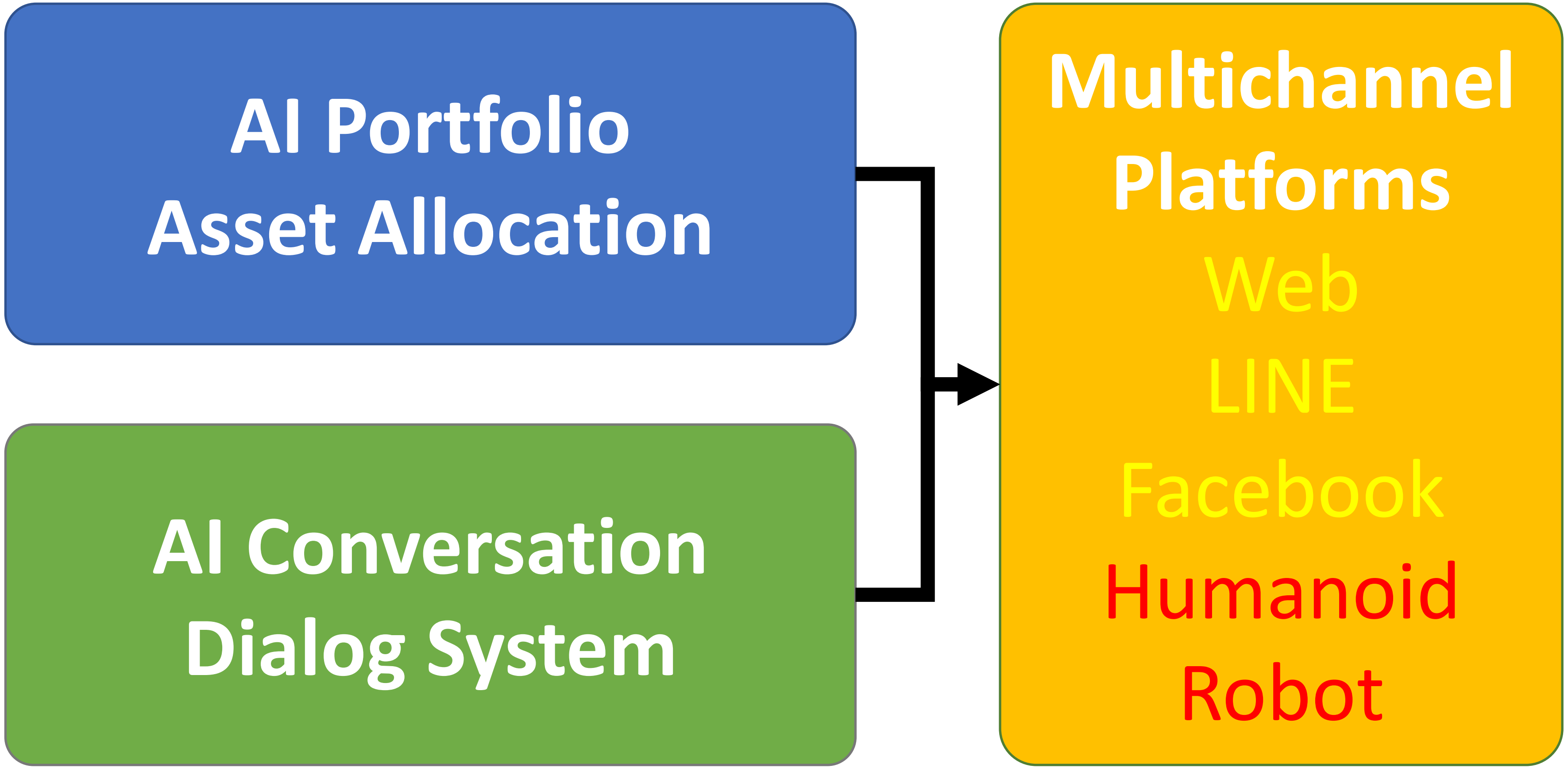
Multi-turn, task-oriented (helpdesk)

Source: https://waseda.app.box.com/v/STC3atNTCIR-14

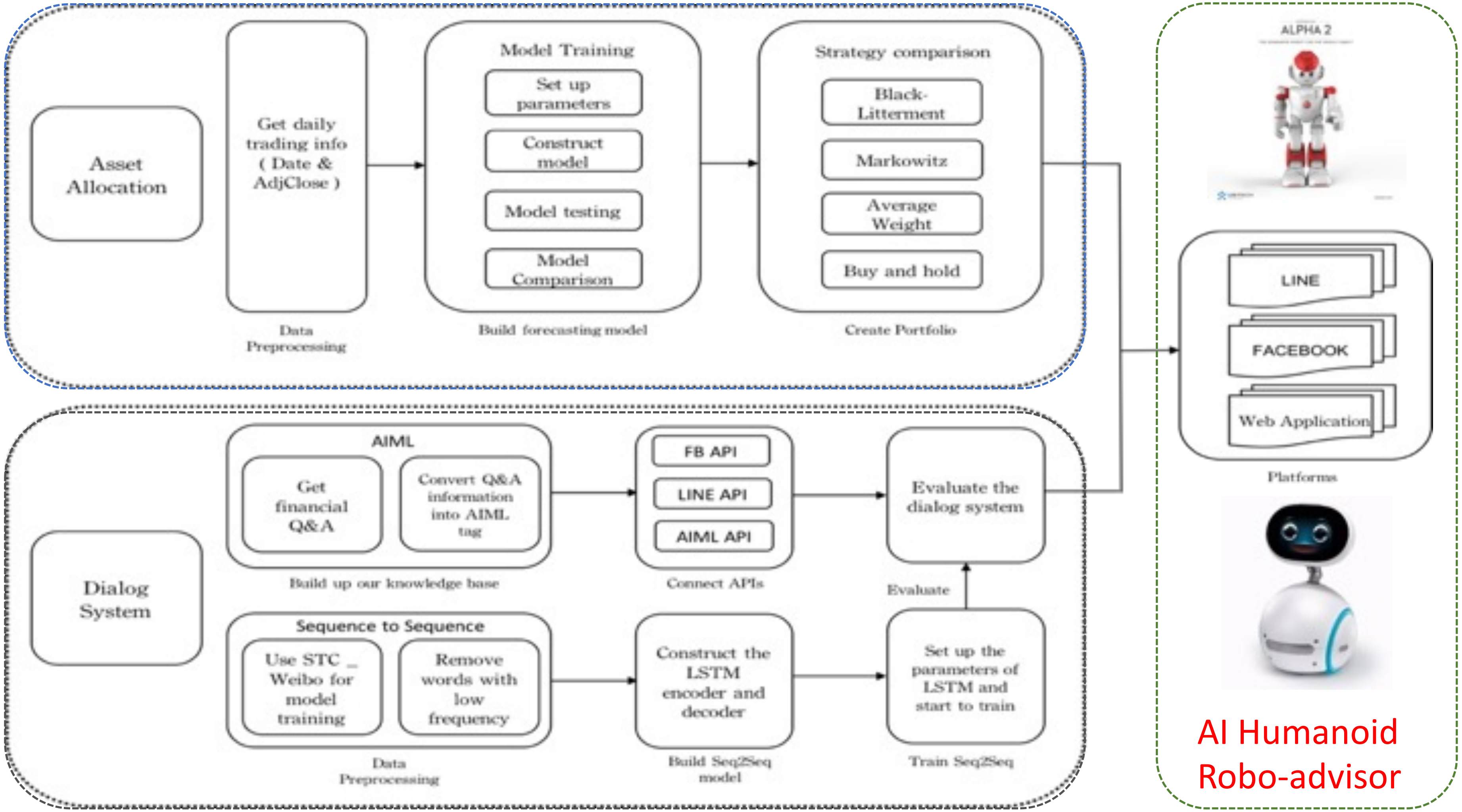
AI Humanoid Robo-Advisor

AI Humanoid Robo-Advisor

for Multi-channel Conversational Commerce



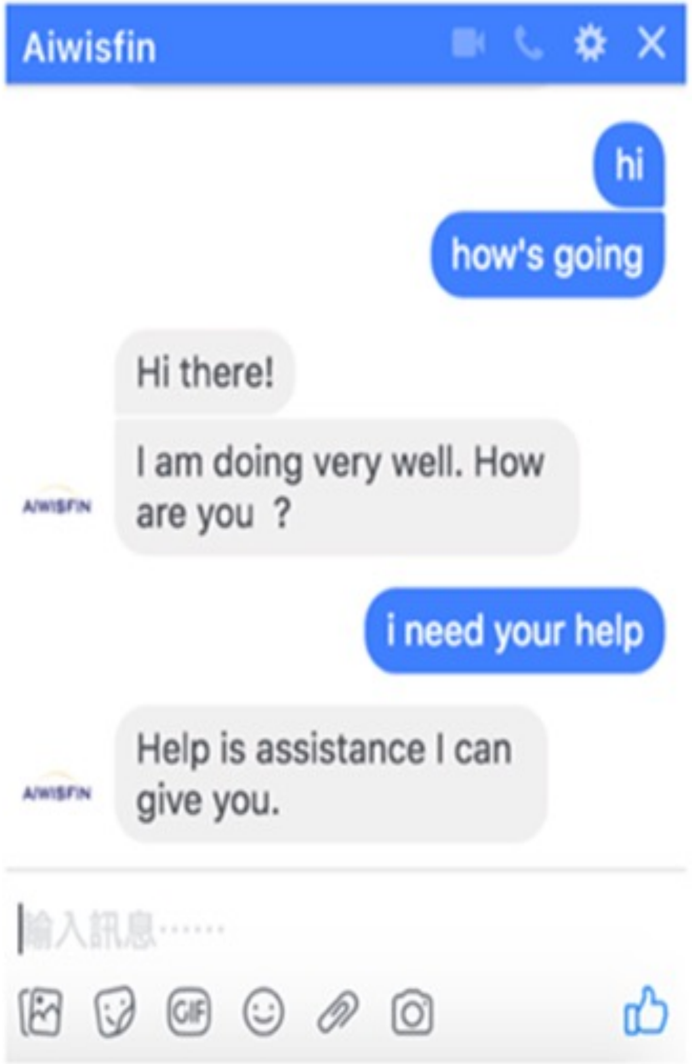
System Architecture of AI Humanoid Robo-Advisor



AI Humanoid Robo-advisor

Conversational Model

(LINE, FB Messenger)



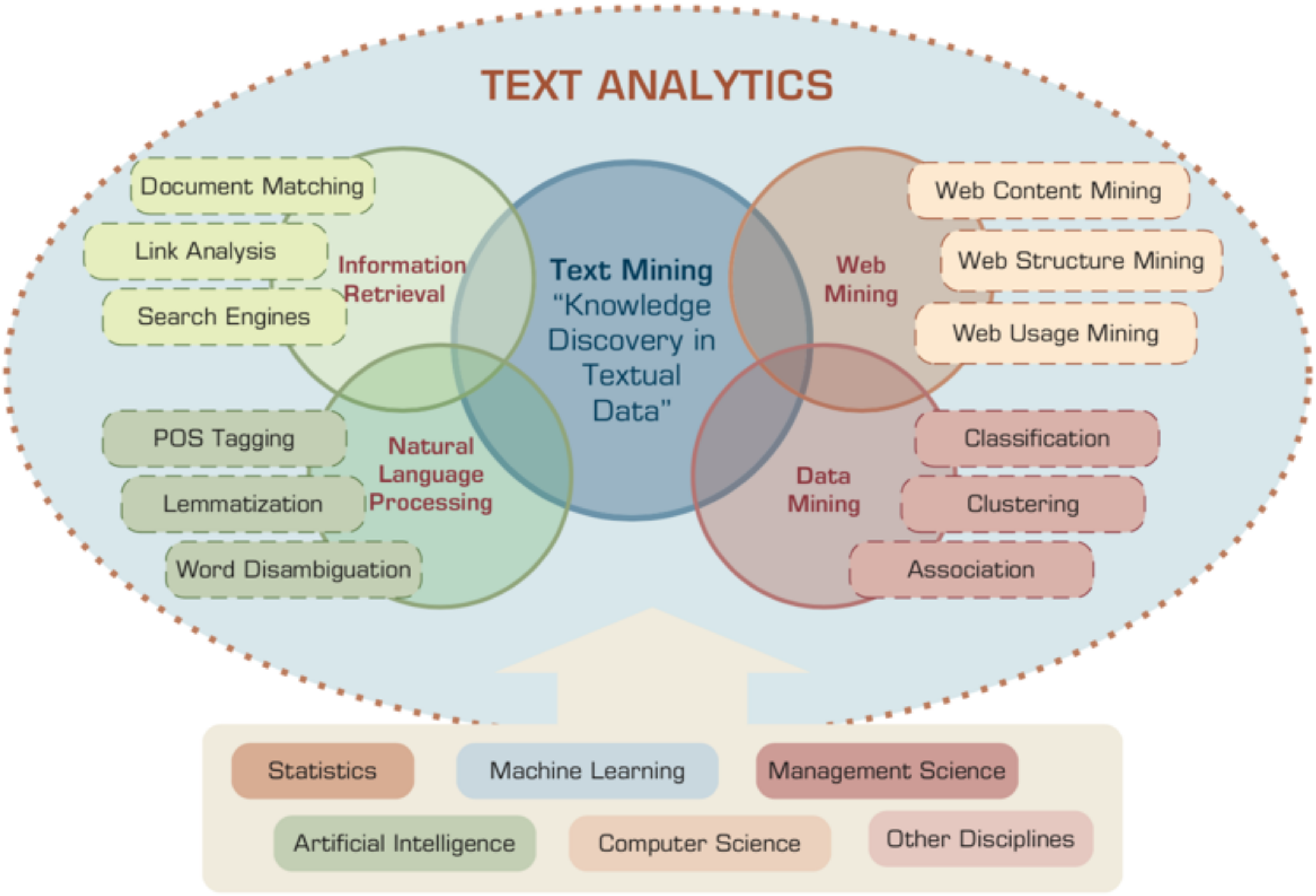
Conversational Robo-Advisor Multichannel UI/UX Robots

ALPHA 2

ZENBO

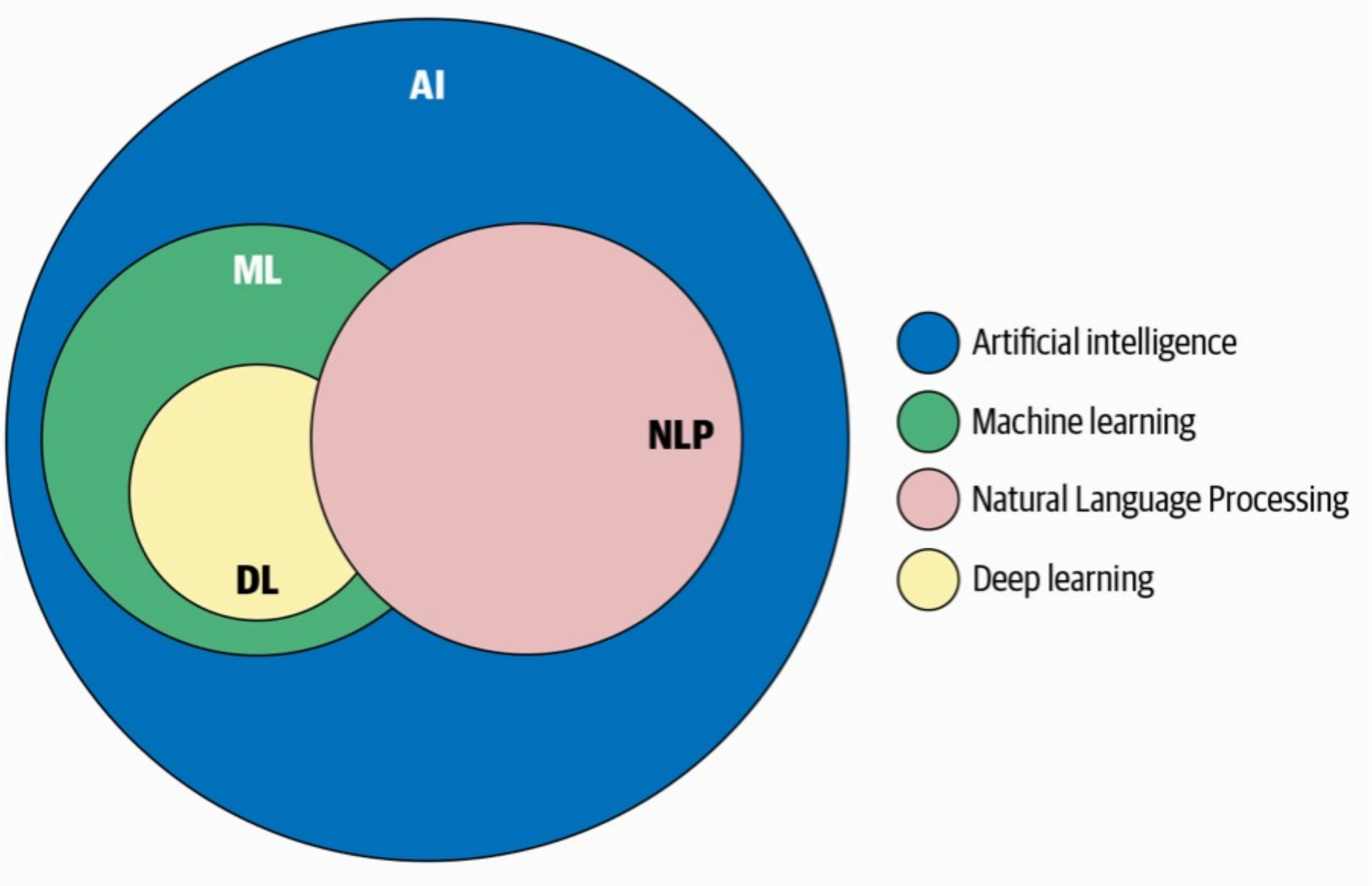
Artificial Intelligence (AI)
for
Text Analytics (TA)

Text Analytics and Text Mining



Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

AI, NLP, ML, DL

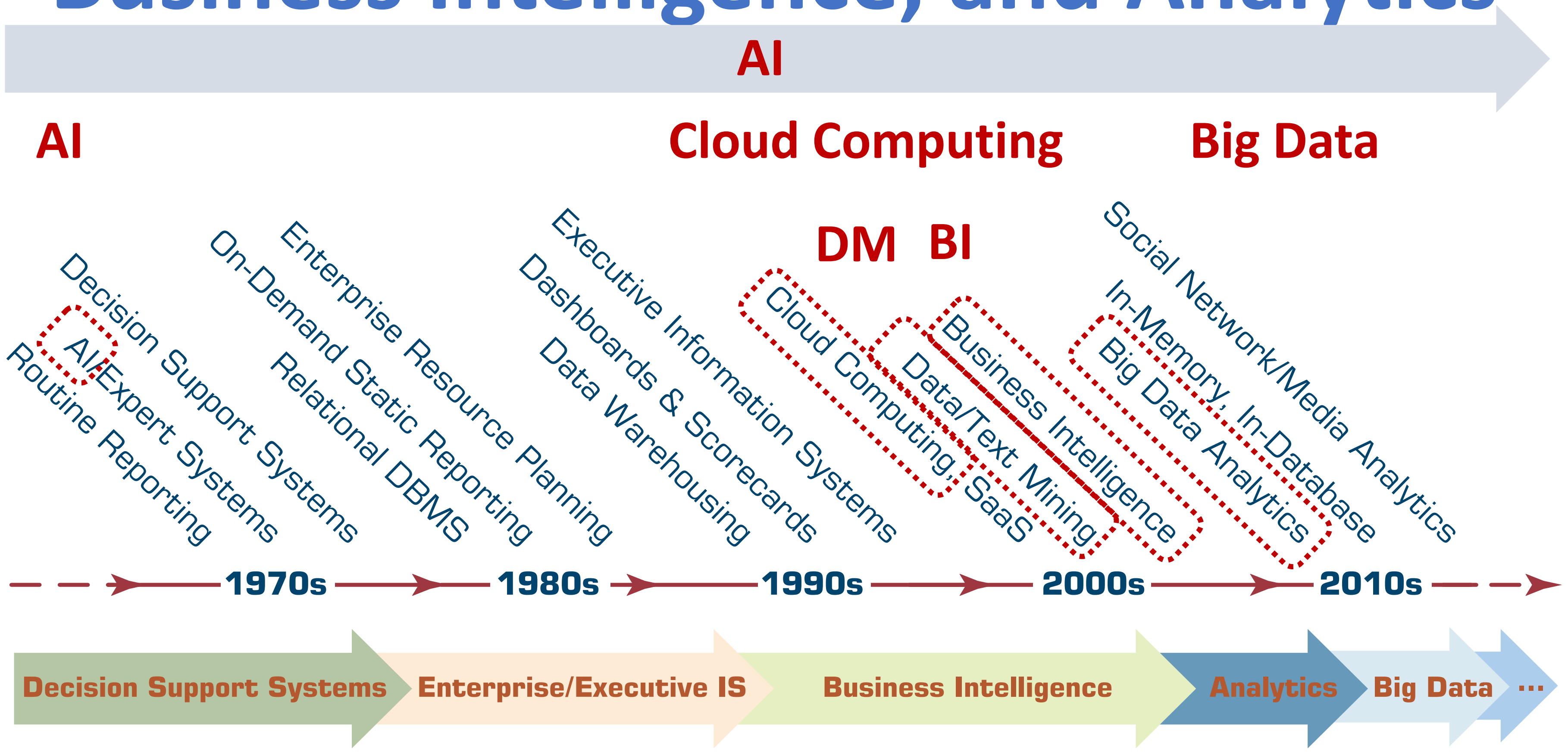


Source: Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta (2020), Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems, O'Reilly Media.

Artificial Intelligence (AI)

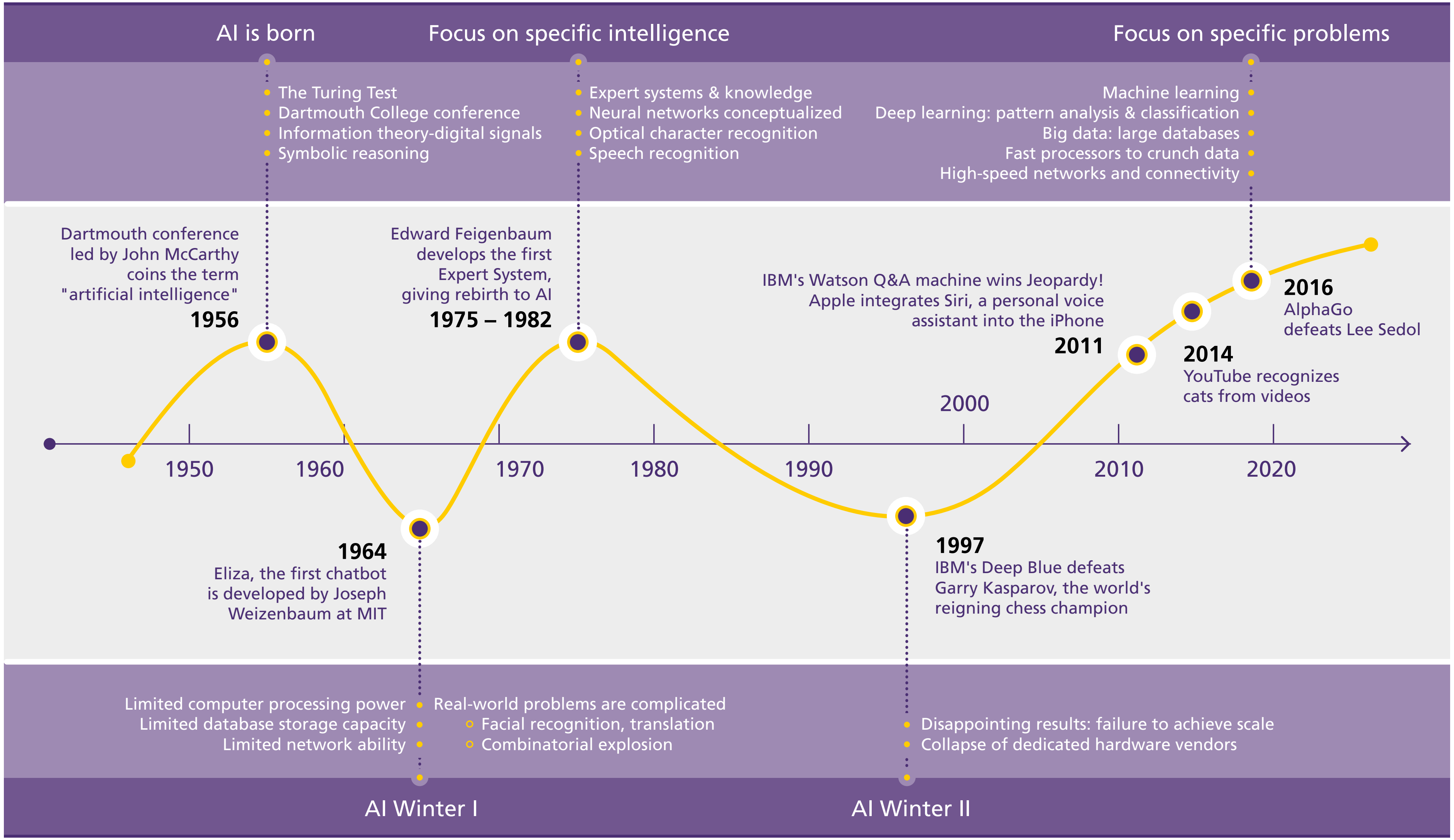
AI, Big Data, Cloud Computing

Evolution of Decision Support, Business Intelligence, and Analytics



Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017),
Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

The Rise of AI



Source: DHL (2018), Artificial Intelligence in Logistics, http://www.globalhha.com/doclib/data/upload/doc_con/5e50c53c5bf67.pdf/

Definition
of
Artificial Intelligence
(A.I.)

Artificial Intelligence

“… the science and

engineering

of

making

intelligent machines”

(John McCarthy, 1955)

Source: https://digitalintelligencetoday.com/artificial-intelligence-defined-useful-list-of-popular-definitions-from-business-and-science/

Artificial Intelligence

“… technology that thinks and acts like humans”

Source: https://digitalintelligencetoday.com/artificial-intelligence-defined-useful-list-of-popular-definitions-from-business-and-science/

Artificial Intelligence

"… intelligence exhibited by machines or software"

Source: https://digitalintelligencetoday.com/artificial-intelligence-defined-useful-list-of-popular-definitions-from-business-and-science/

4 Approaches of AI

Thinking Humanly	Thinking Rationally
Acting Humanly	Acting Rationally

Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

4 Approaches of AI

2. Thinking Humanly: The Cognitive Modeling Approach	3. Thinking Rationally: The "Laws of Thought" Approach
1. Acting Humanly: The Turing Test Approach (1950)	4. Acting Rationally: The Rational Agent Approach

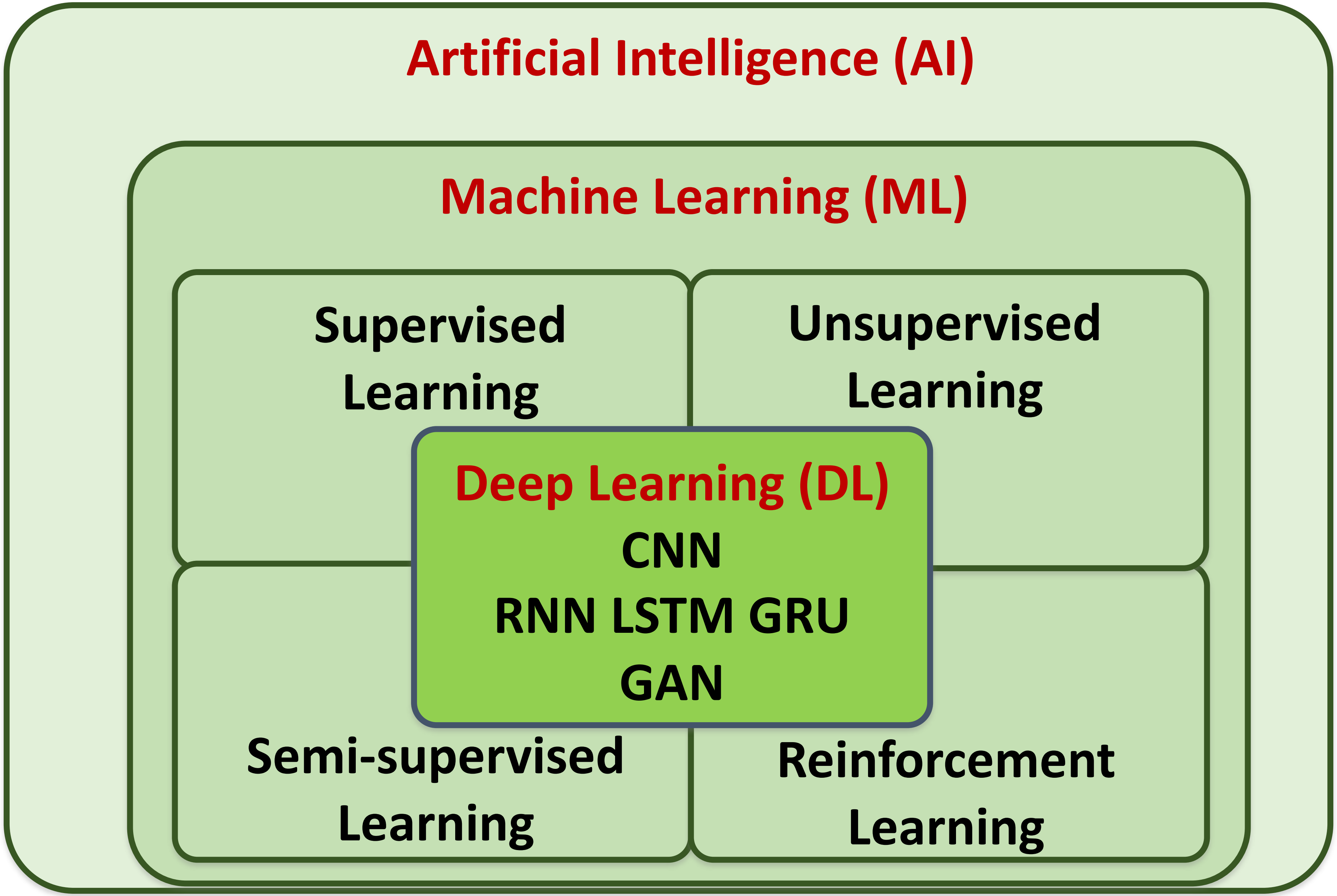
Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

AI Acting Humanly: The Turing Test Approach
(Alan Turing, 1950)

- Knowledge Representation
- Automated Reasoning
- Machine Learning (ML)
  - Deep Learning (DL)
- Computer Vision (Image, Video)
- Natural Language Processing (NLP)
- Robotics

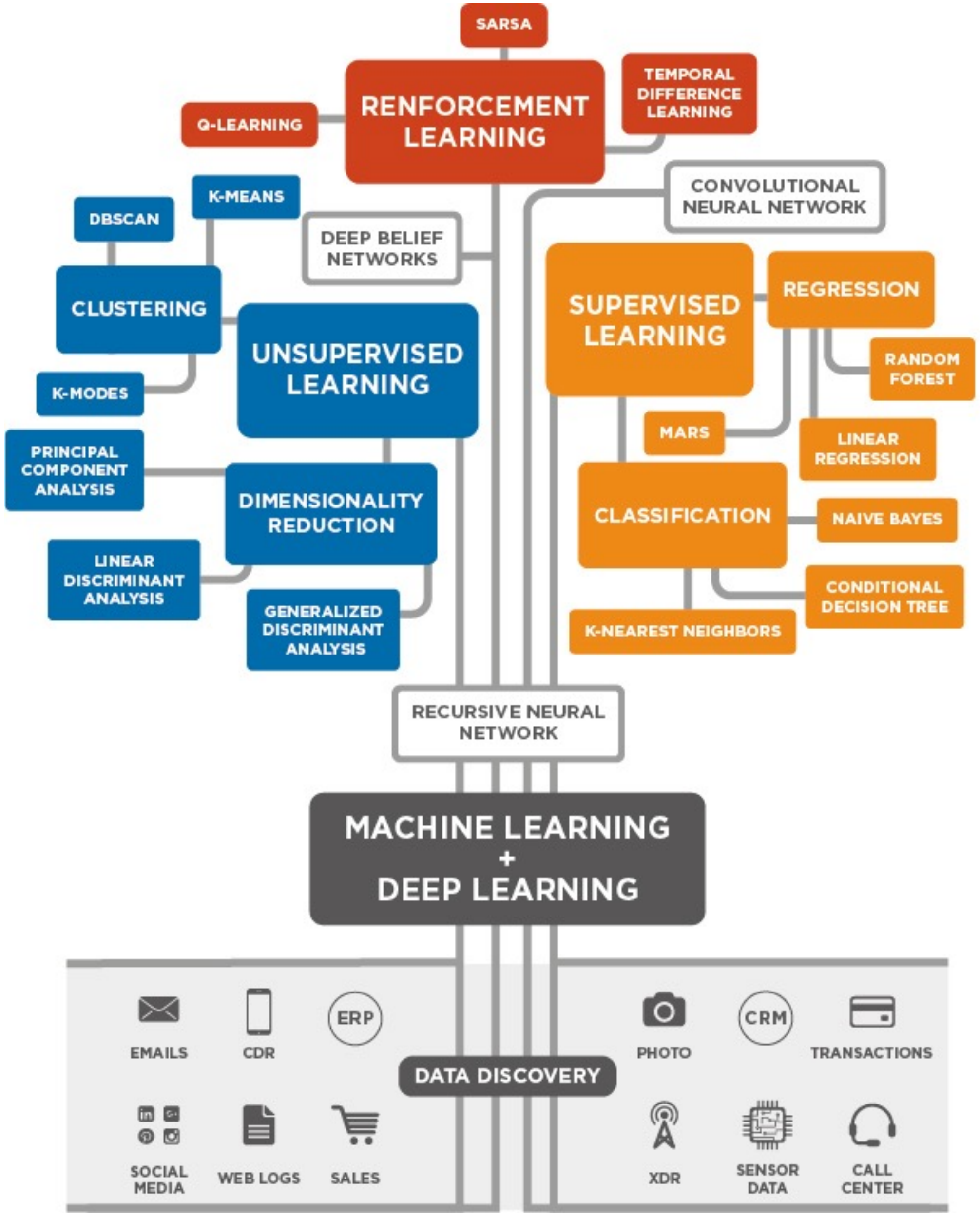
Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

AI, ML, DL



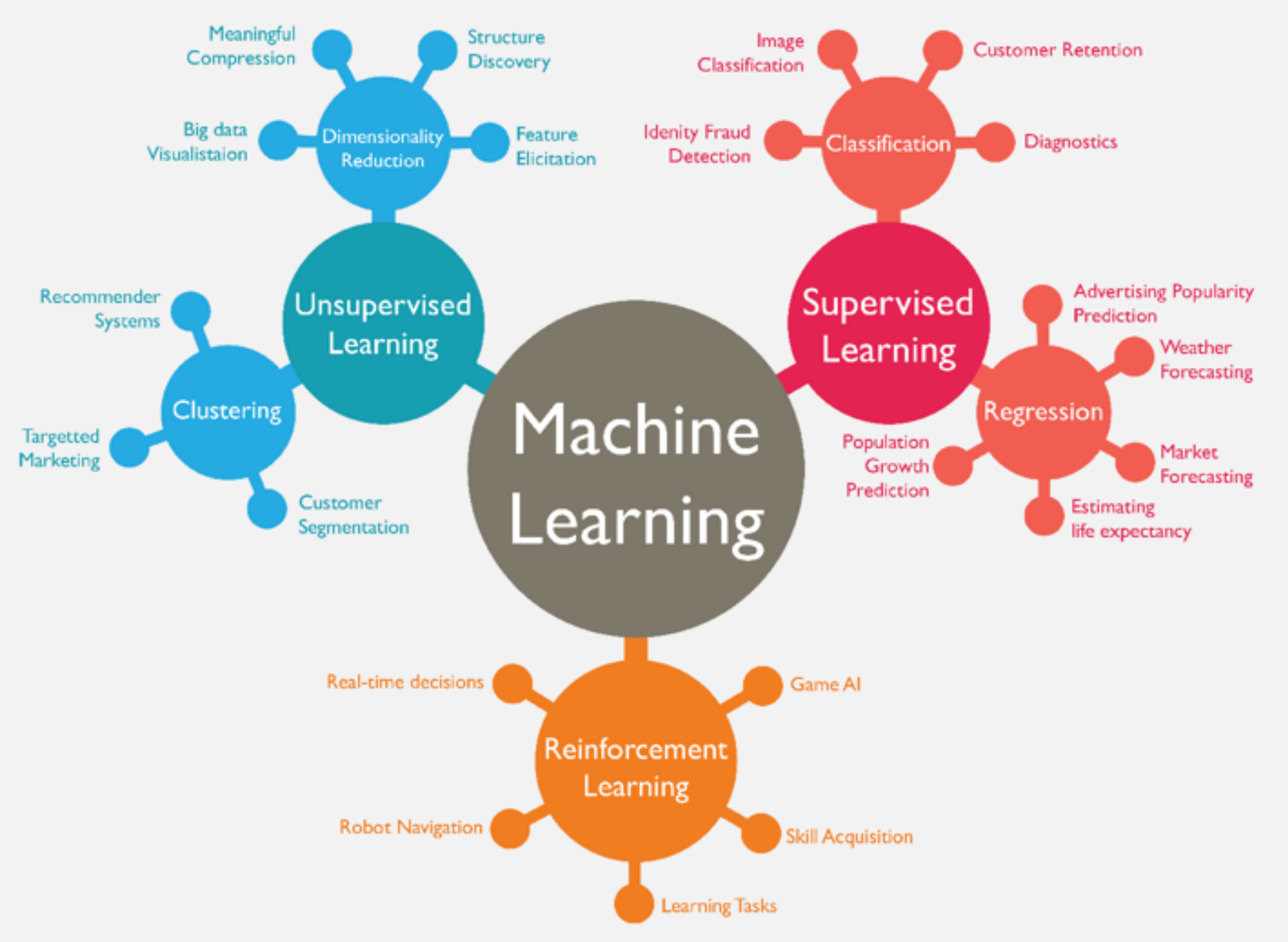
Source: https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/deep_learning.html

3 Machine Learning Algorithms



Source: Enrico Galimberti, http://blogs.teradata.com/data-points/tree-machine-learning-algorithms/

Machine Learning (ML)



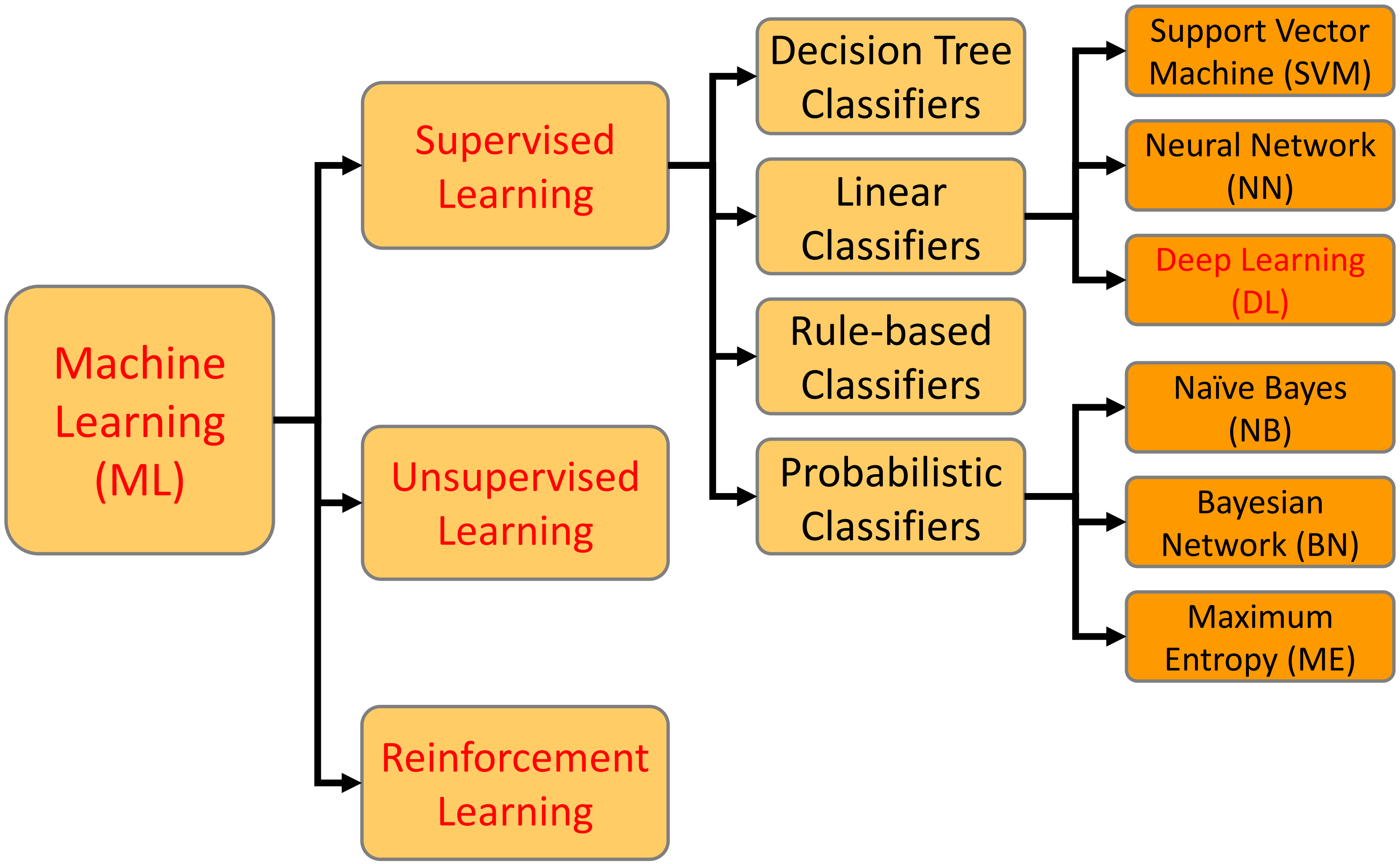
Source: https://www.mactores.com/services/aws-big-data-machine-learning-cognitive-services/

Machine Learning Models

- Deep Learning
- Kernel
- Association rules
- Ensemble
- Decision tree
- Dimensionality reduction
- Clustering
- Regression Analysis
- Bayesian
- Instance based

Source: Sunila Gollapudi (2016), Practical Machine Learning, Packt Publishing

Machine Learning (ML) / Deep Learning (DL)



Source: Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.

Text Analytics and Text Mining

Text Analytics (TA)

Text Analytics

- Text Analytics = Information Retrieval + Information Extraction + Data Mining + Web Mining
- Text Analytics = Information Retrieval + Text Mining

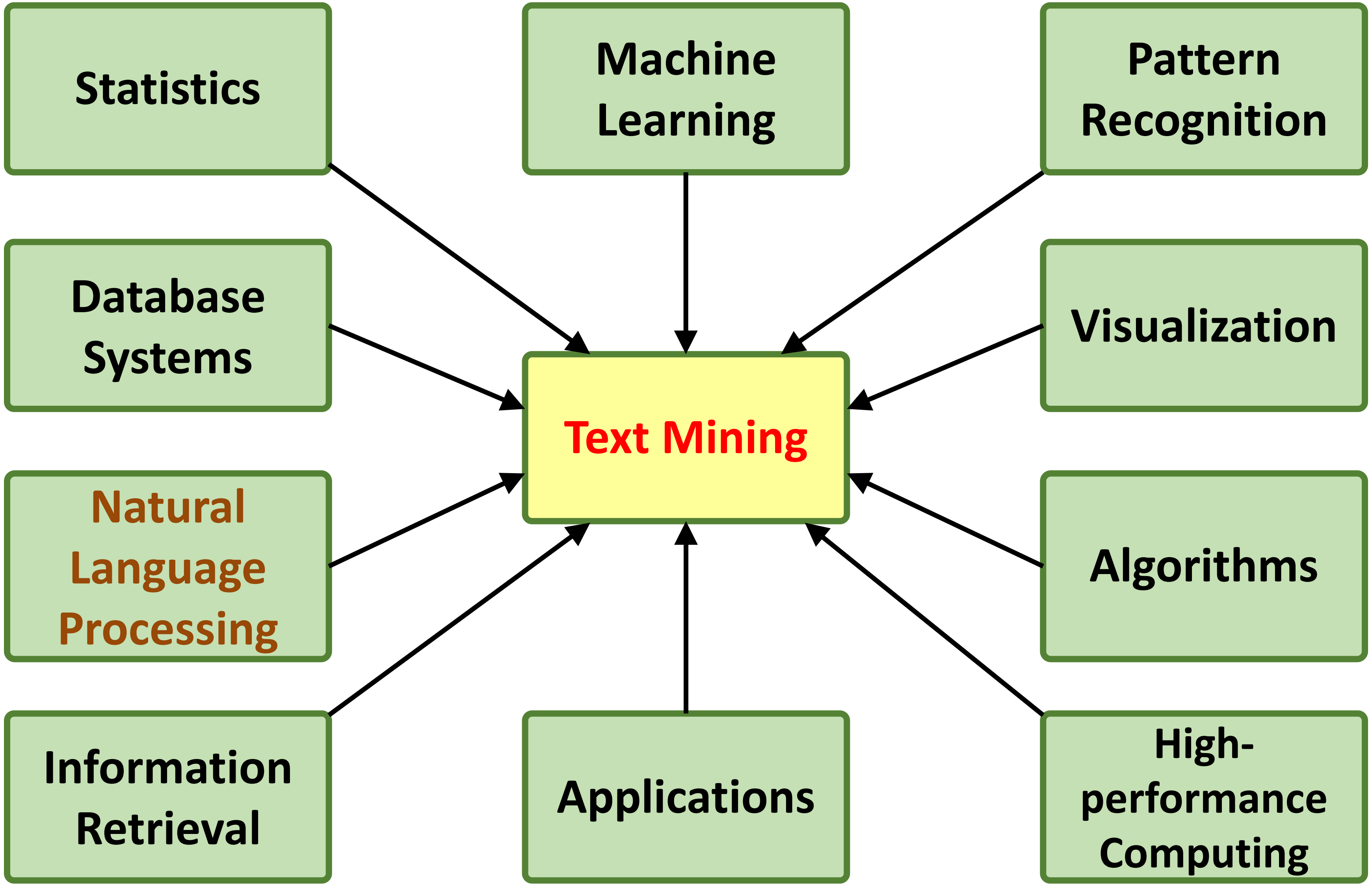
Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Text Mining

- Text Data Mining
- Knowledge Discovery in Textual Databases

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Text Mining Technologies



Application Areas of Text Mining

- Information extraction
- Topic tracking
- Summarization
- Categorization
- Clustering
- Concept linking
- Question answering

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Text Mining Technologies

Text Mining
(TM)

Natural Language Processing
(NLP)

Text mining

Text Data Mining

Intelligent Text Analysis

Knowledge-Discovery in Text (KDT)

Source: Vishal Gupta and Gurpreet S. Lehal (2009), "A survey of text mining techniques and applications," Journal of emerging technologies in web intelligence, vol. 1, no. 1, pp. 60-76.

Text Mining
(text data mining)

the process of
deriving
high-quality information
from text

http://en.wikipedia.org/wiki/Text_mining

Text Mining:

the process of extracting interesting and non-trivial information and knowledge from unstructured text.

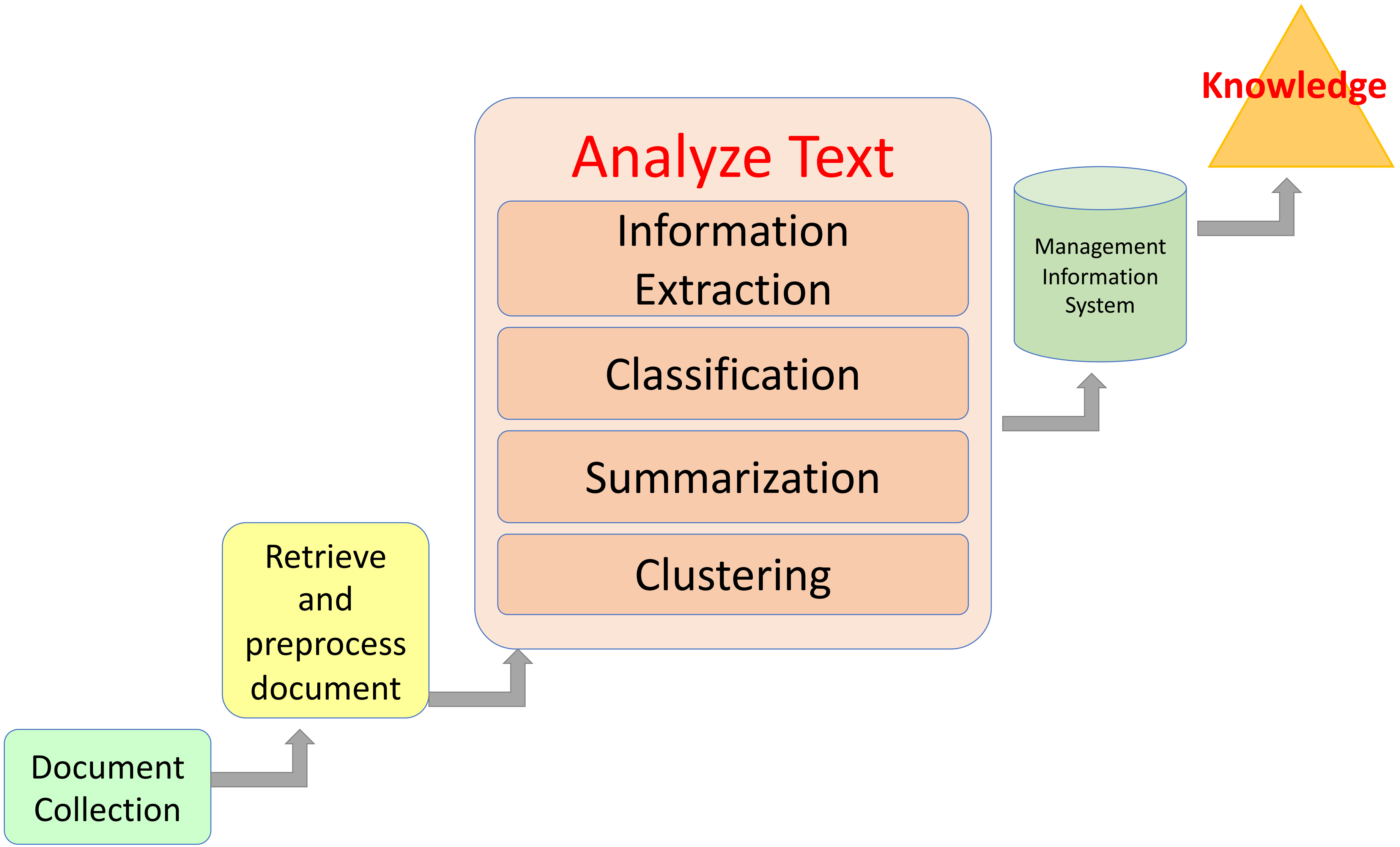
Source: Vishal Gupta and Gurpreet S. Lehal (2009), "A survey of text mining techniques and applications," Journal of emerging technologies in web intelligence, vol. 1, no. 1, pp. 60-76.

Text Mining:

discovery by computer of new, previously unknown information, by automatically extracting information from different written resources.

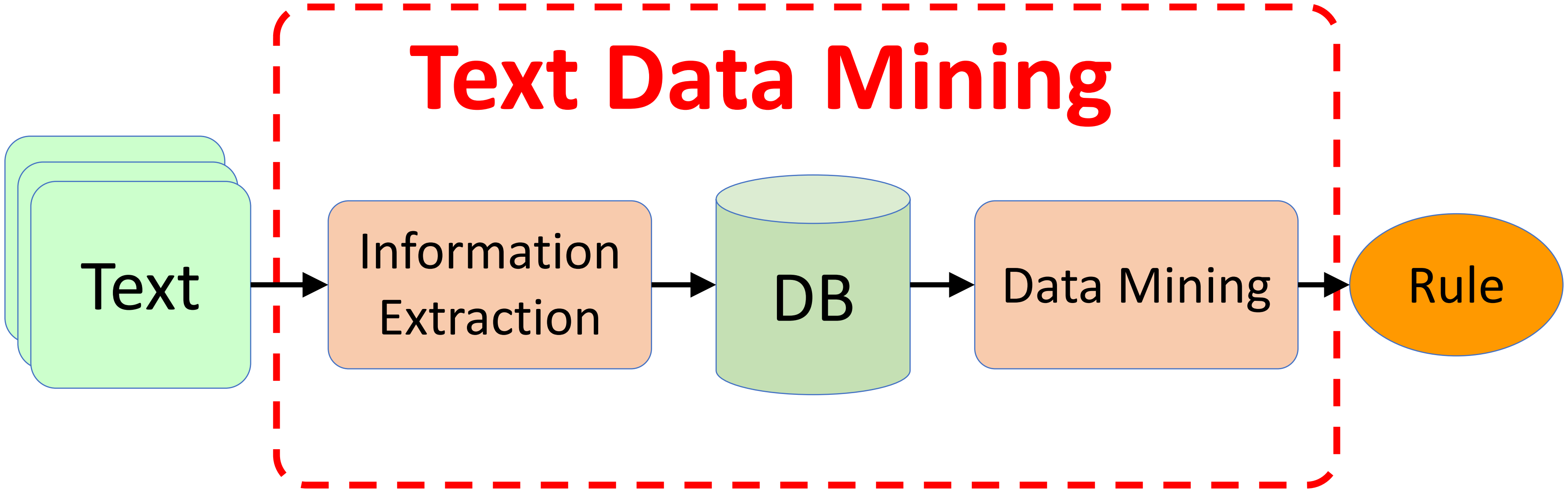
Source: Vishal Gupta and Gurpreet S. Lehal (2009), "A survey of text mining techniques and applications," Journal of emerging technologies in web intelligence, vol. 1, no. 1, pp. 60-76.

An example of Text Mining



Source: Vishal Gupta and Gurpreet S. Lehal (2009), "A survey of text mining techniques and applications," Journal of emerging technologies in web intelligence, vol. 1, no. 1, pp. 60-76.

Overview of Information Extraction based Text Mining Framework



Source: Vishal Gupta and Gurpreet S. Lehal (2009), "A survey of text mining techniques and applications," Journal of emerging technologies in web intelligence, vol. 1, no. 1, pp. 60-76.

Natural Language Processing (NLP)

- **Natural language processing (NLP)** is an **important component of text mining** and is a **subfield of artificial intelligence and computational linguistics**.

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Natural Language Processing with Transformers:

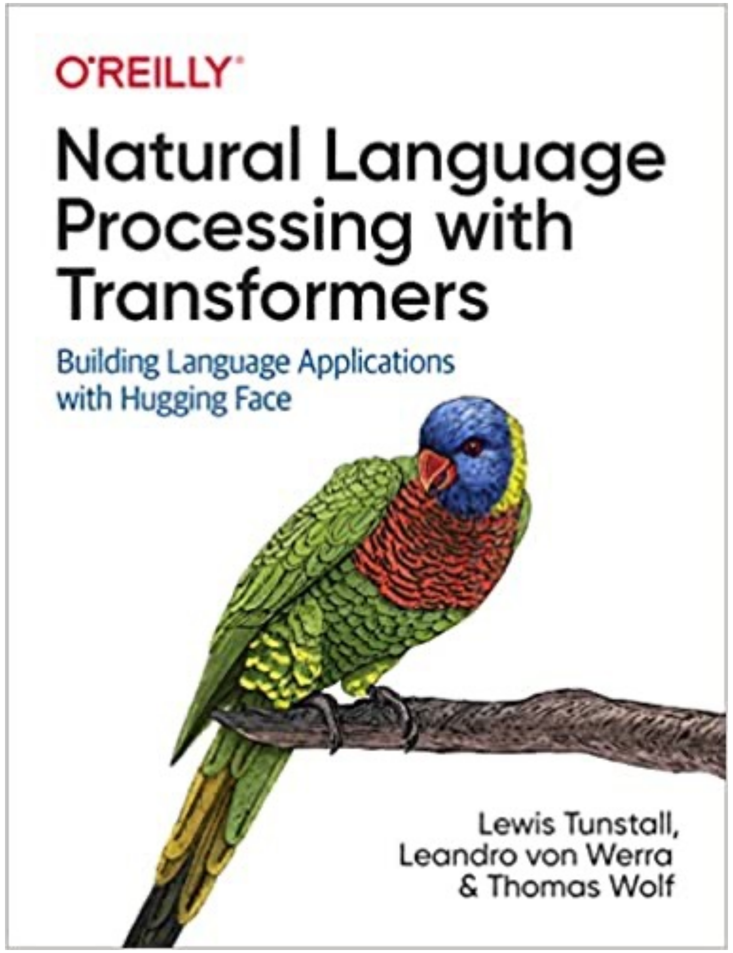
Building Language Applications with Hugging Face

Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022),

Natural Language Processing with Transformers:

Building Language Applications with Hugging Face,

O'Reilly Media.



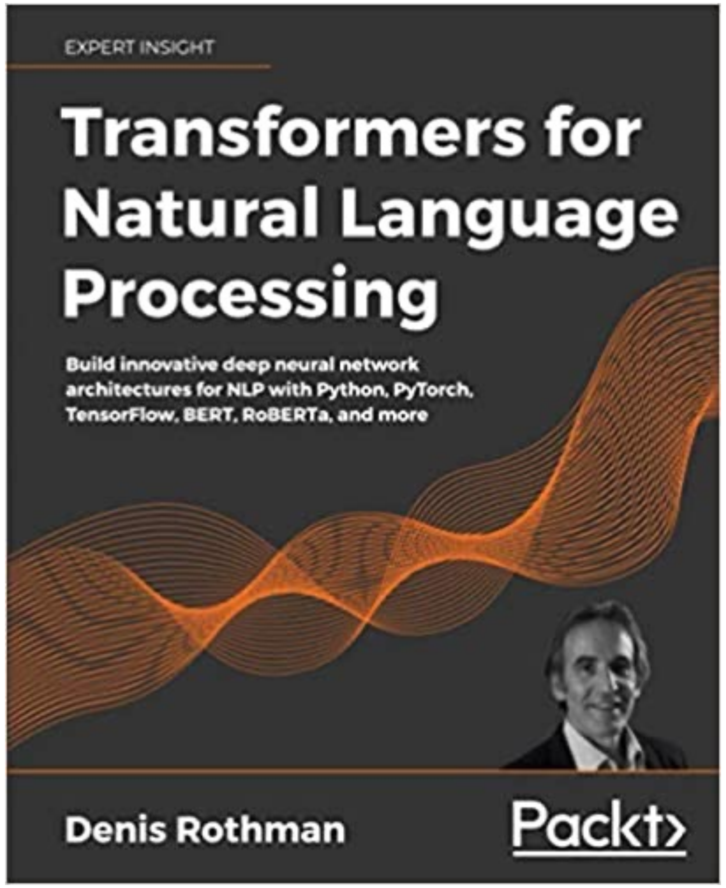
Source: https://www.amazon.com/Natural-Language-Processing-Transformers-Applications/dp/1098103246

Denis Rothman (2021),

Transformers for Natural Language Processing:

Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more,

Packt Publishing.



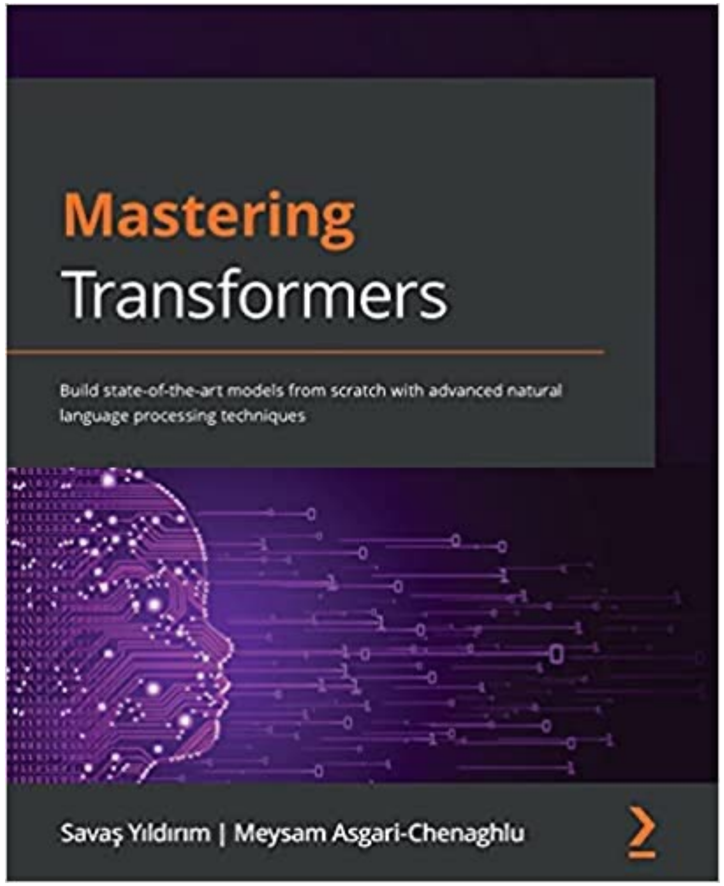
Source: https://www.amazon.com/Transformers-Natural-Language-Processing-architectures/dp/1800565798

Savaş Yıldırım and Meysam Asgari-Chenaghlu (2021),

Mastering Transformers:

Build state-of-the-art models from scratch with advanced natural language processing techniques,

Packt Publishing.



Source: https://www.amazon.com/Mastering-Transformers-state-art-processing/dp/1801077657/

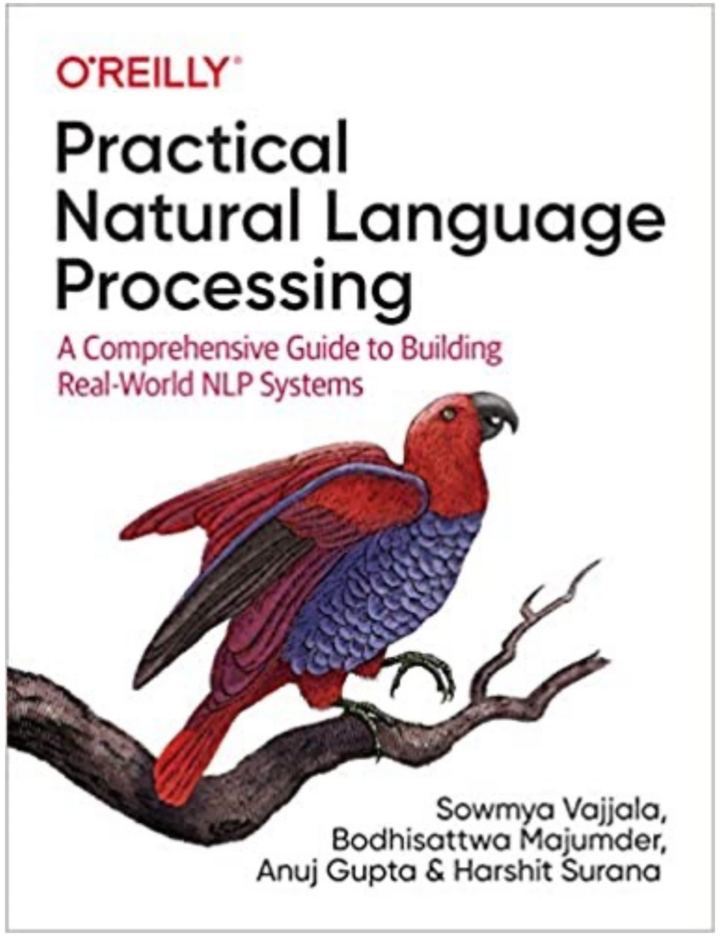
Practical
Natural Language Processing

Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta (2020),

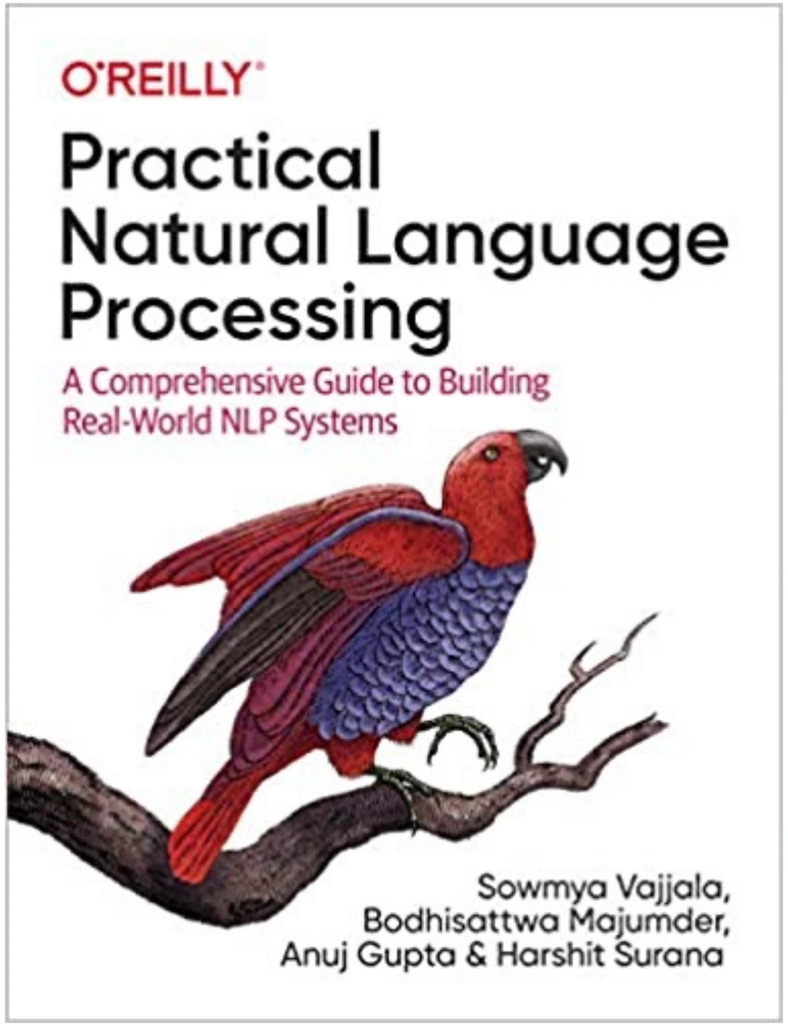
Practical Natural Language Processing:

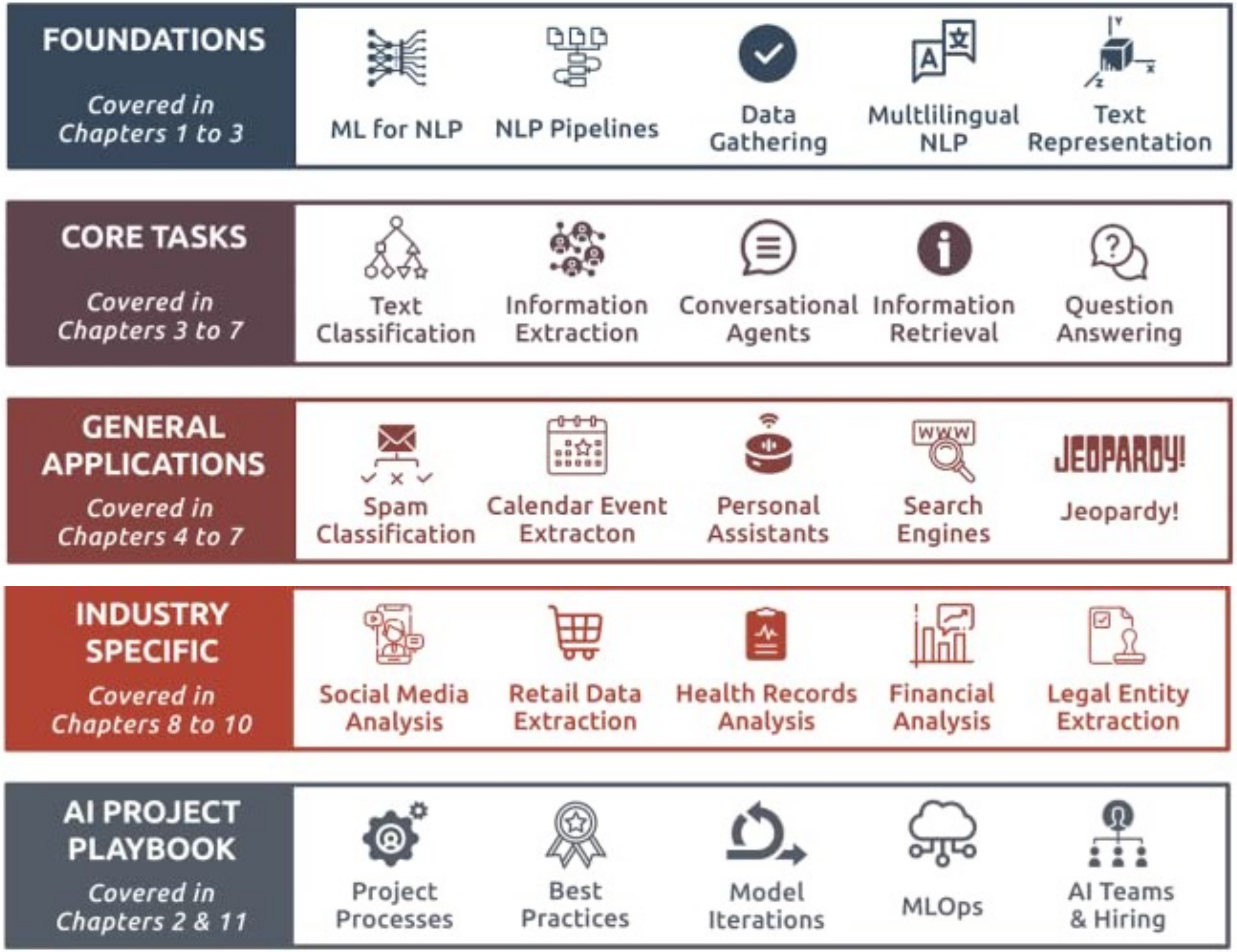
A Comprehensive Guide to Building Real-World NLP Systems,

O'Reilly Media.



Source: https://www.amazon.com/Practical-Natural-Language-Processing-Pragmatic/dp/1492054054

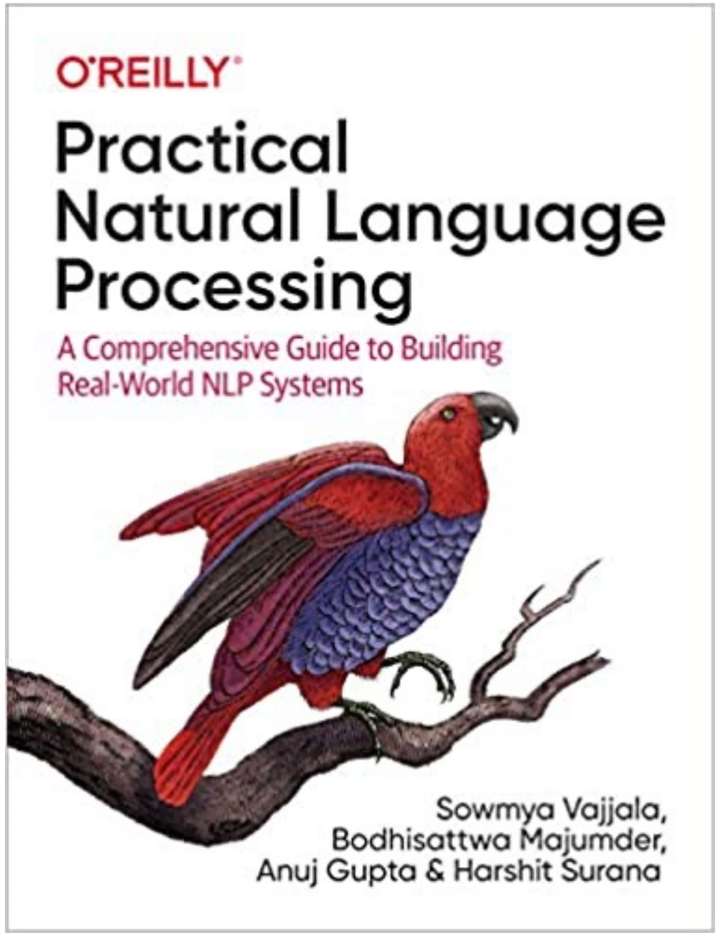


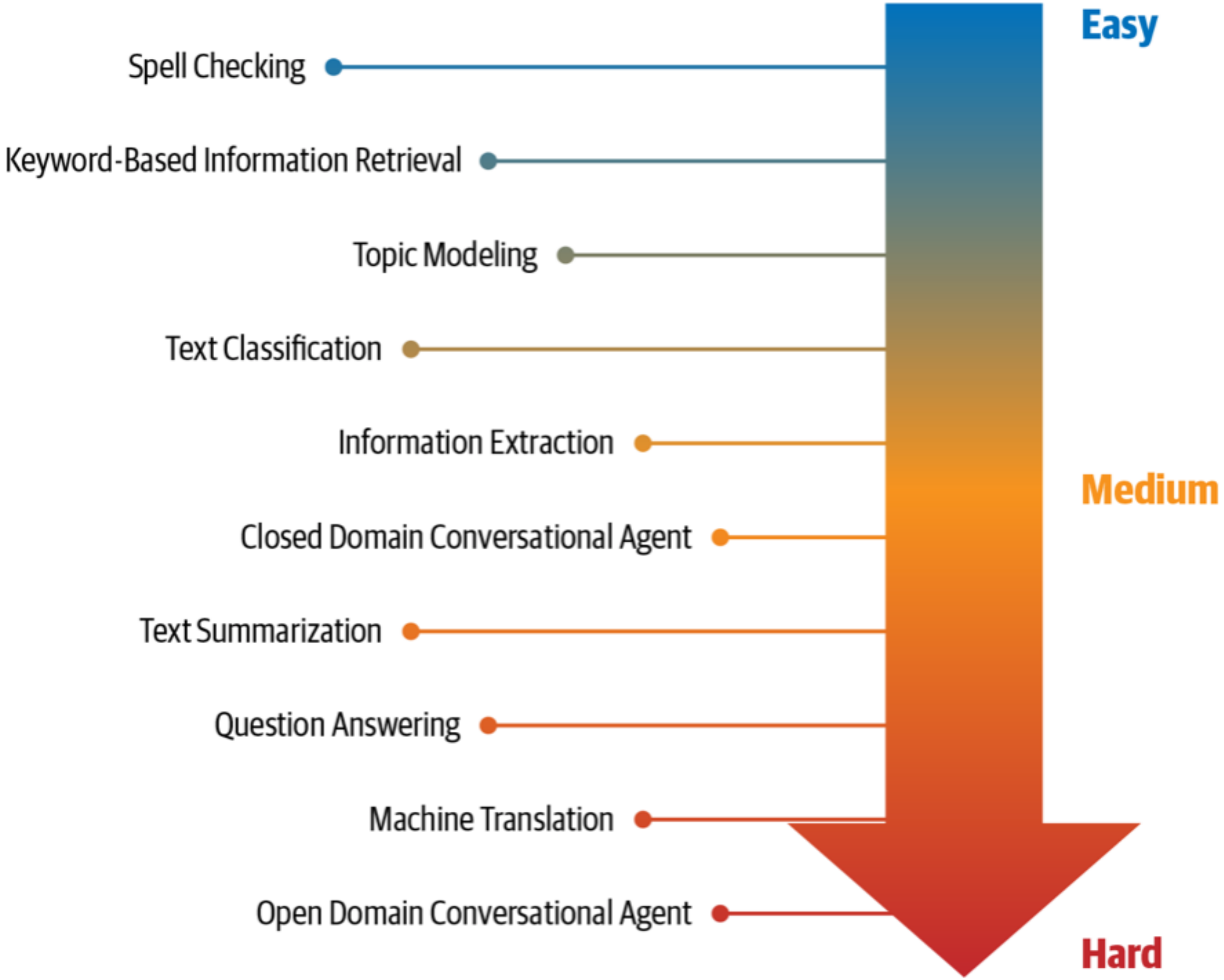


Source: Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta (2020), Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems, O'Reilly Media.

Source: https://www.amazon.com/Practical-Natural-Language-Processing-Pragmatic/dp/1492054054

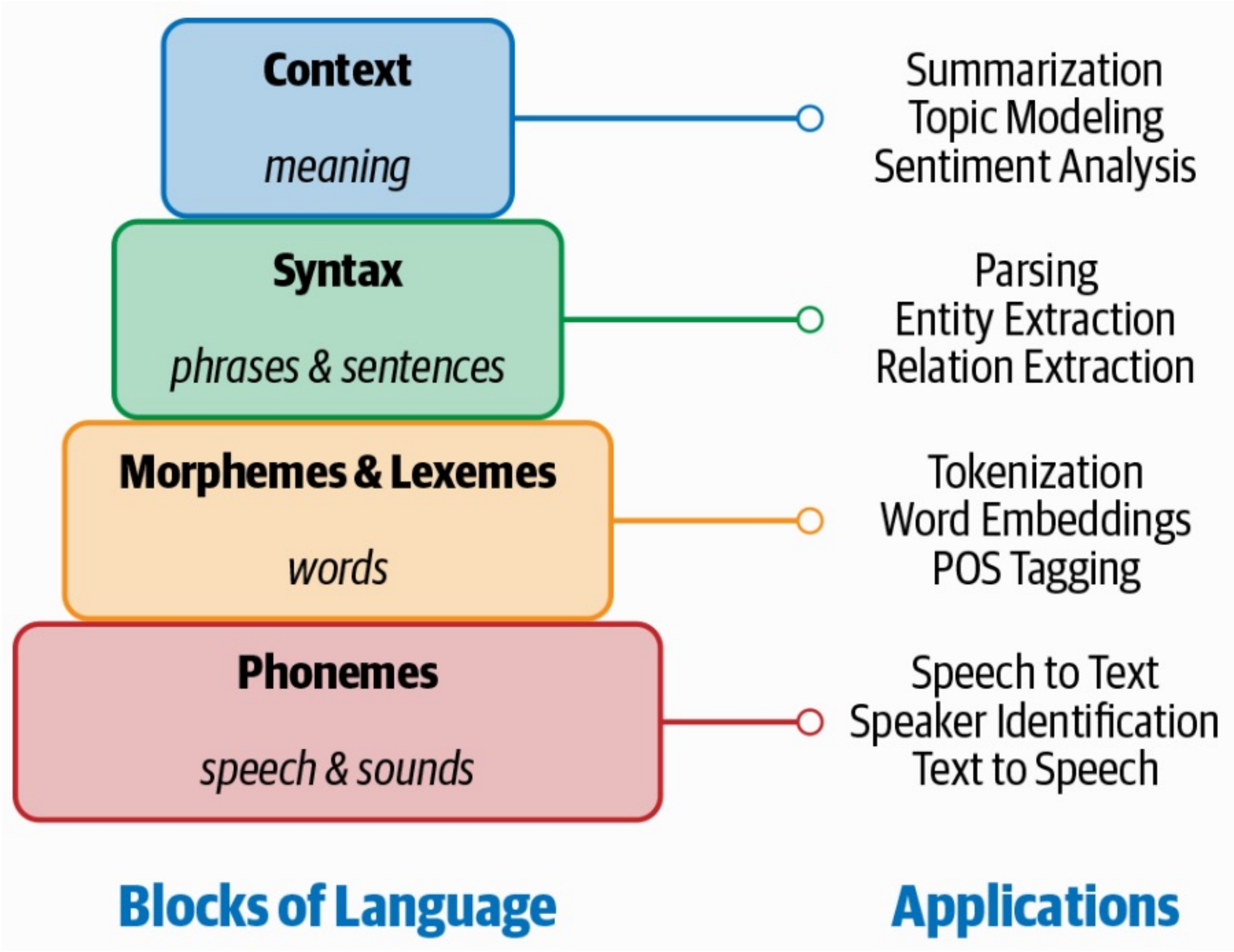
NLP Tasks





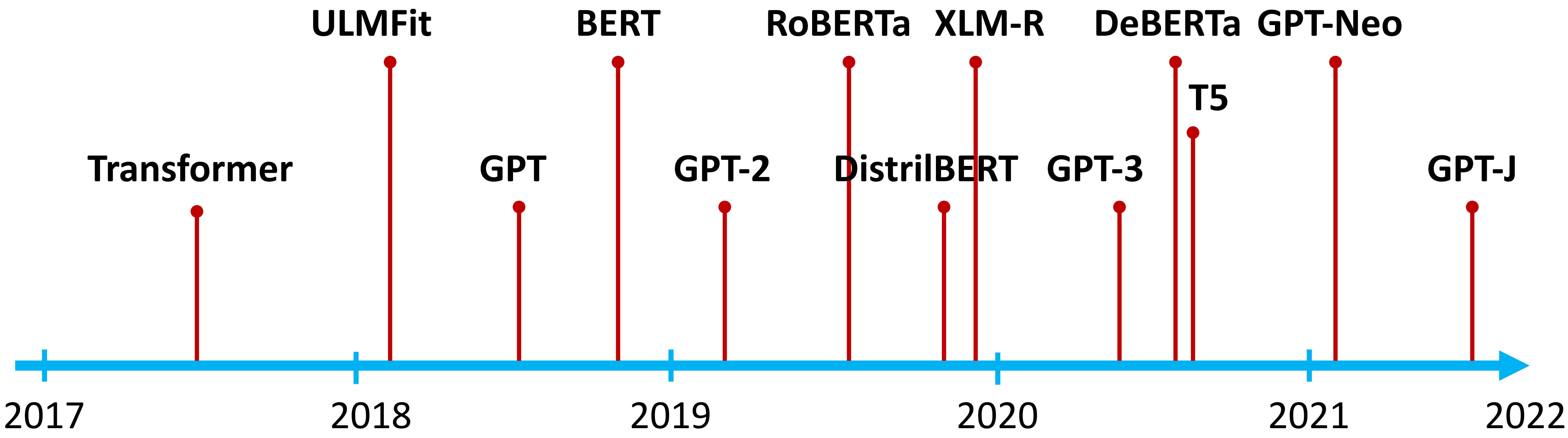
Source: Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta (2020), Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems, O'Reilly Media.

Building Blocks of Language and Applications



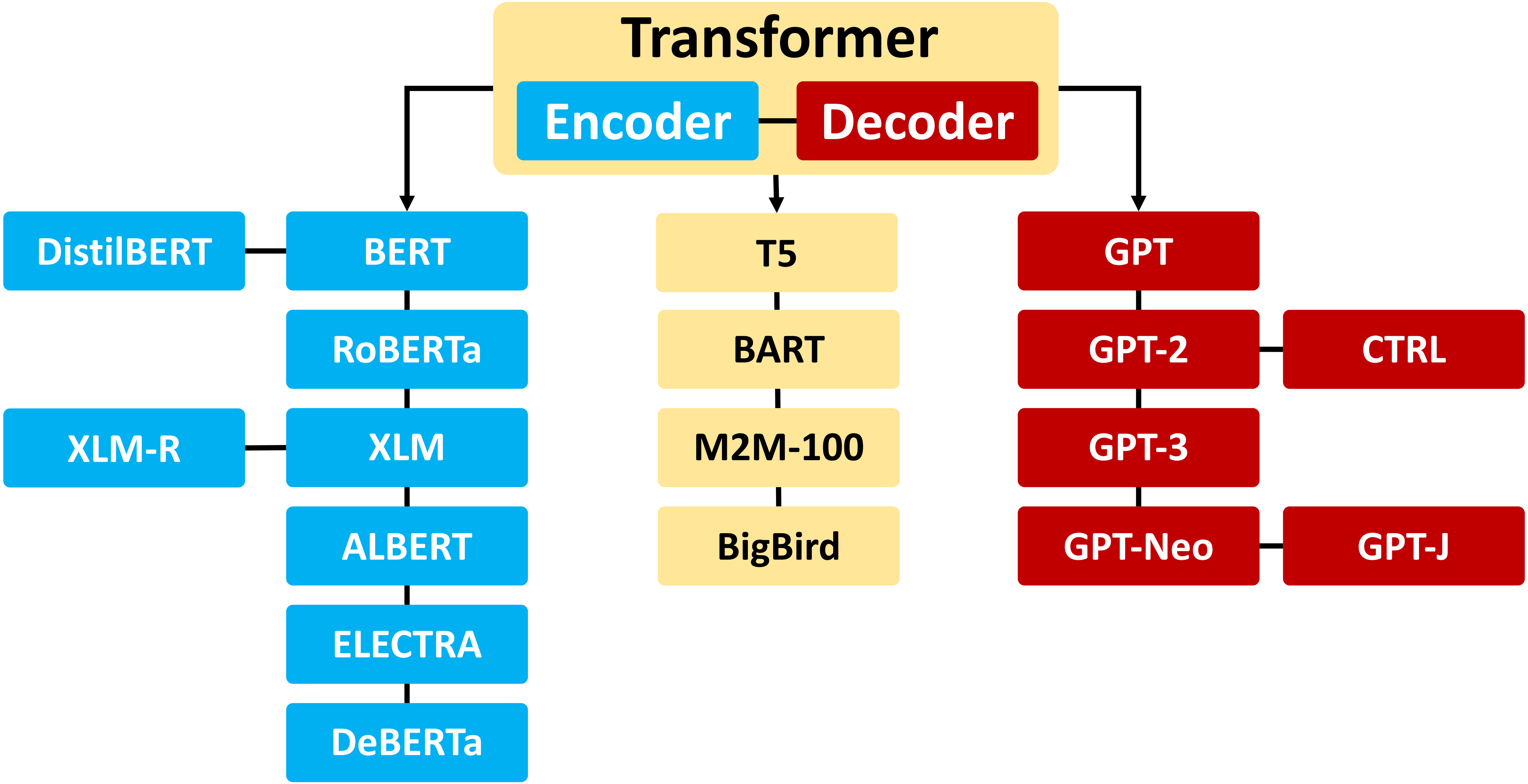
Source: Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta (2020), Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems, O'Reilly Media.

The Transformers Timeline



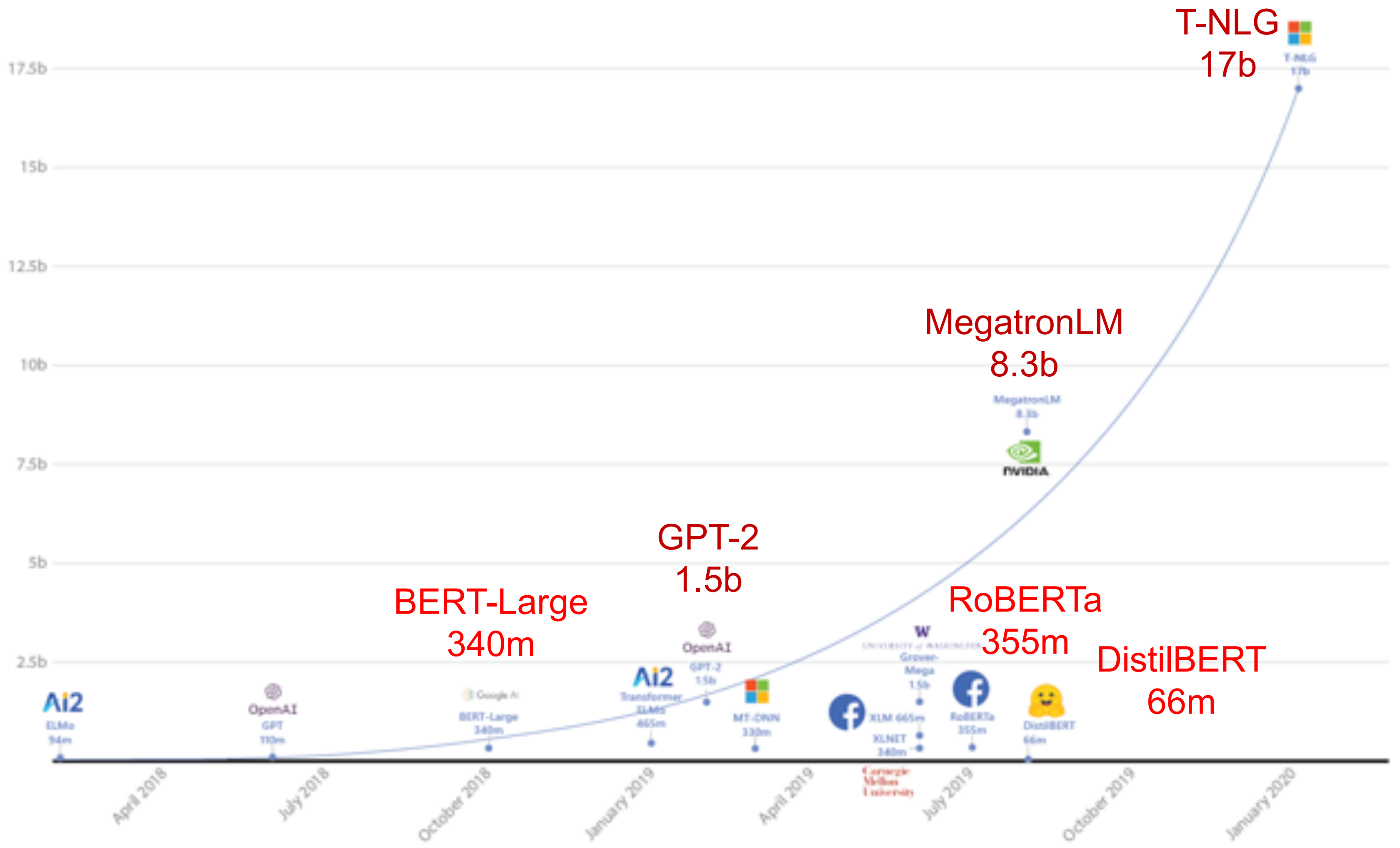
Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.

Transformer Models



Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.

Transformers Pre-trained Language Model

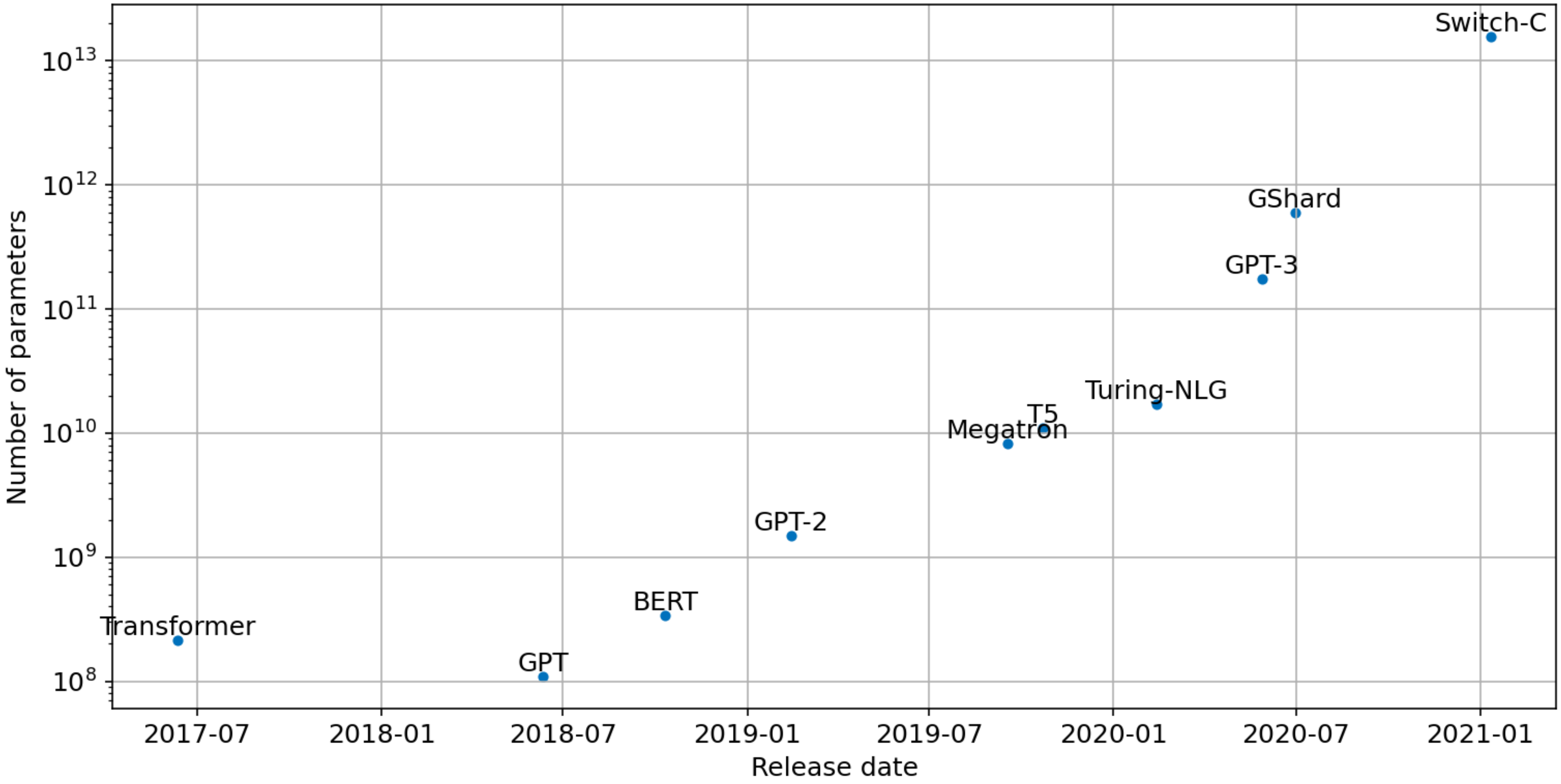


90+ Models

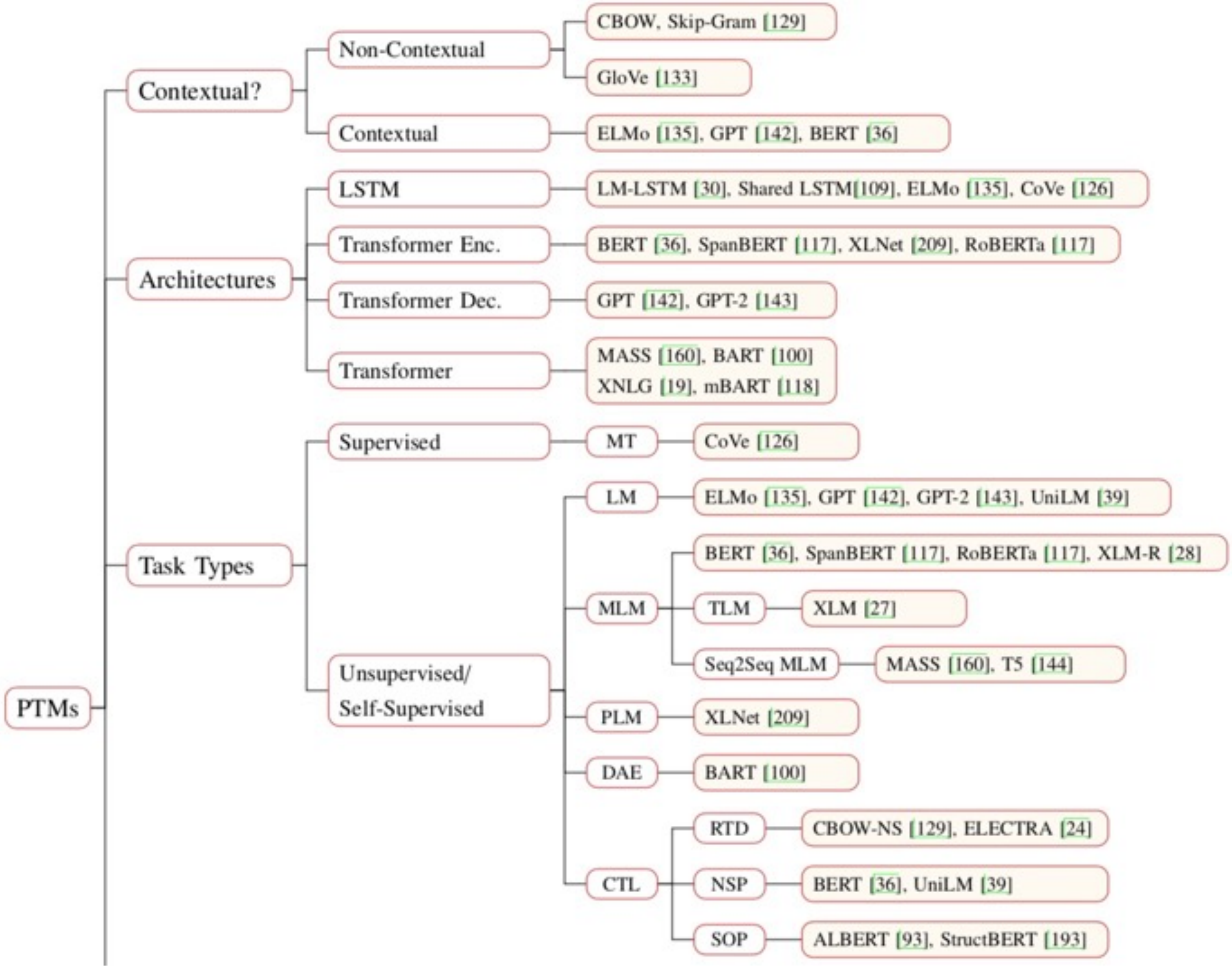
2018 2019 2020 2021 2022

Source: https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/

Scaling Transformers

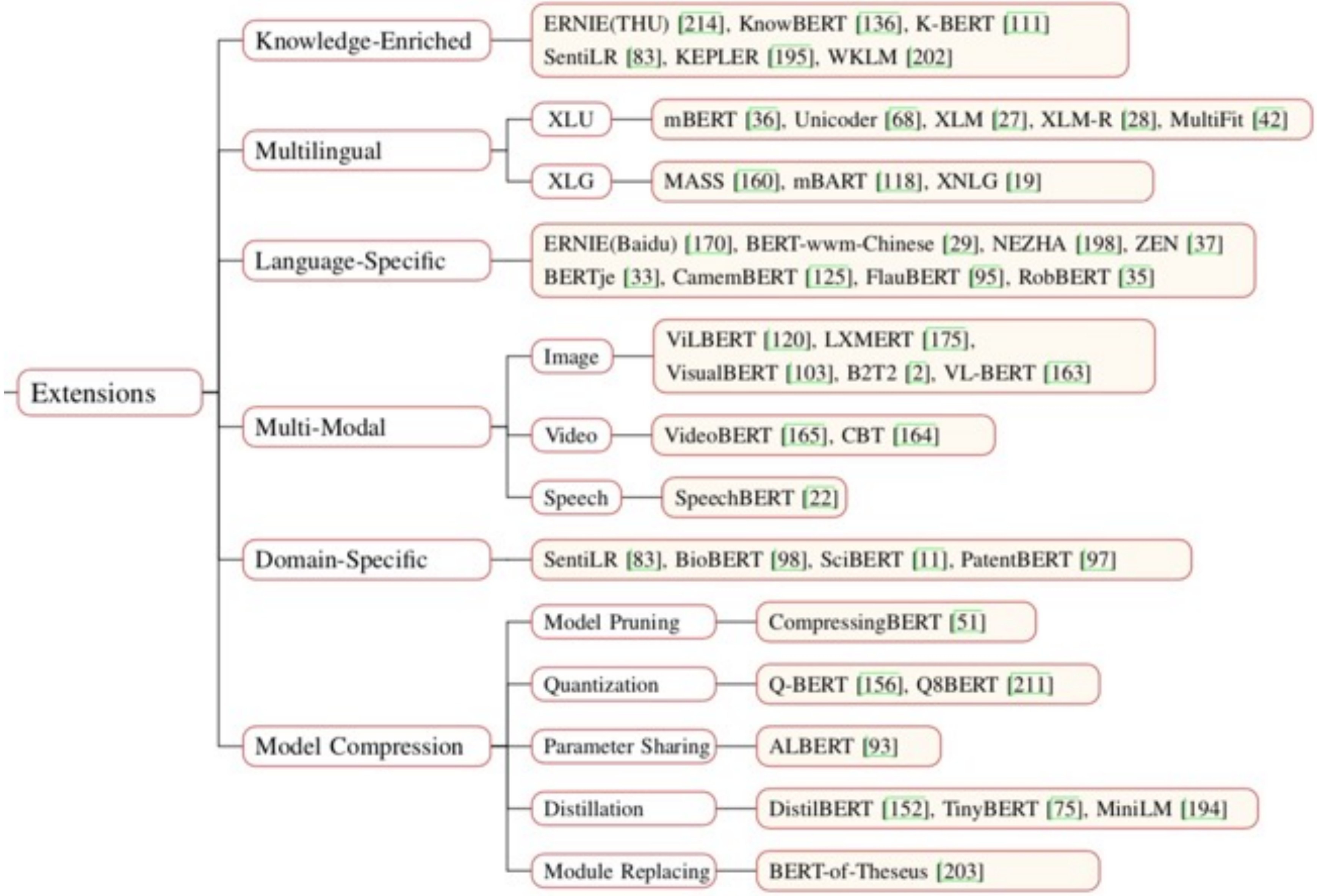


Pre-trained Models (PTM)



Source: Qiu, Xipeng, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. "Pre-trained Models for Natural Language Processing: A Survey." arXiv preprint arXiv:2003.08271 (2020).

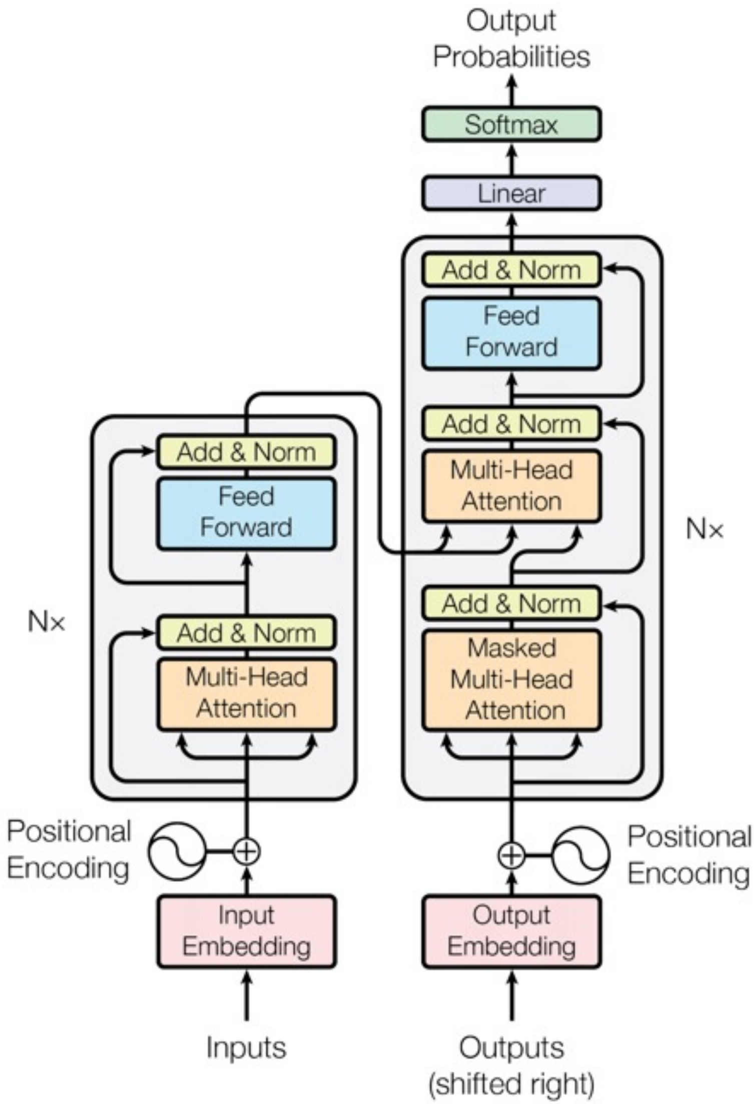
Pre-trained Models (PTM)



Source: Qiu, Xipeng, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. "Pre-trained Models for Natural Language Processing: A Survey." arXiv preprint arXiv:2003.08271 (2020).

Transformer (Attention is All You Need)

(Vaswani et al., 2017)

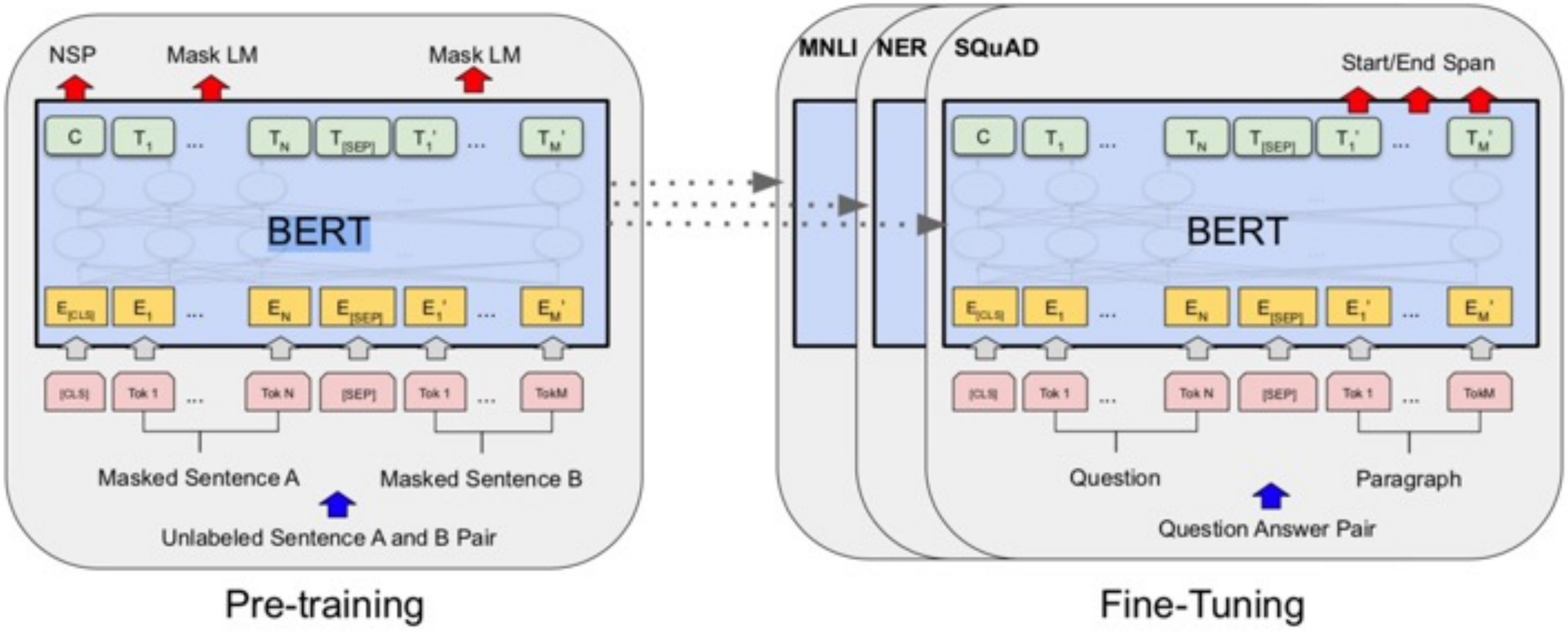


Source: Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In *Advances in neural information processing systems*, pp. 5998-6008. 2017.

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

BERT (Bidirectional Encoder Representations from Transformers)

Overall pre-training and fine-tuning procedures for BERT

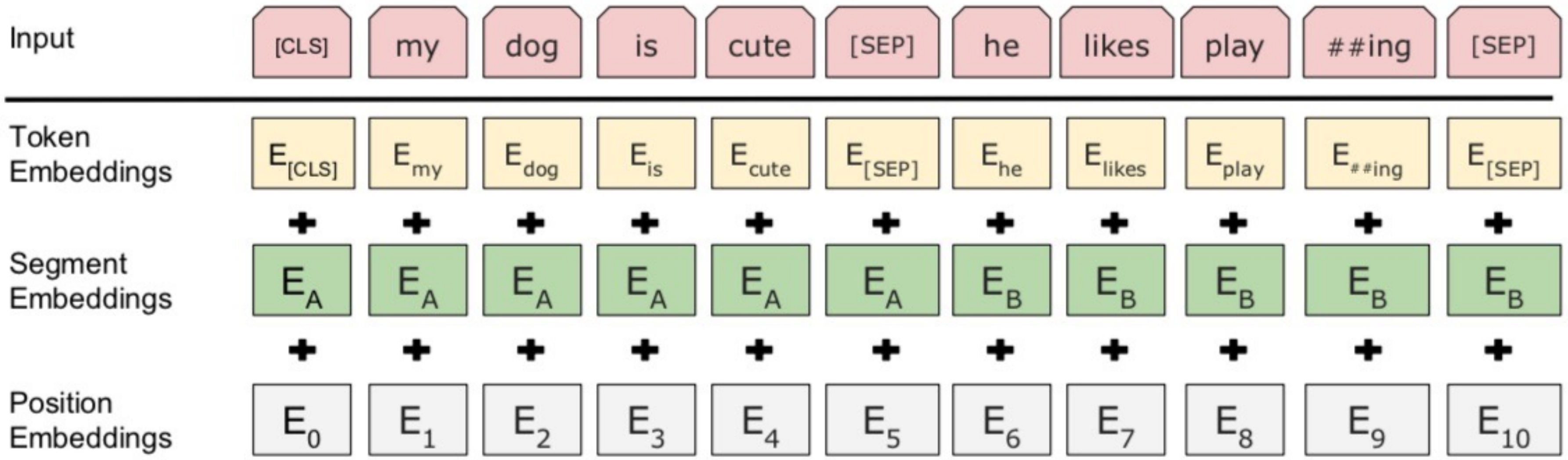


Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

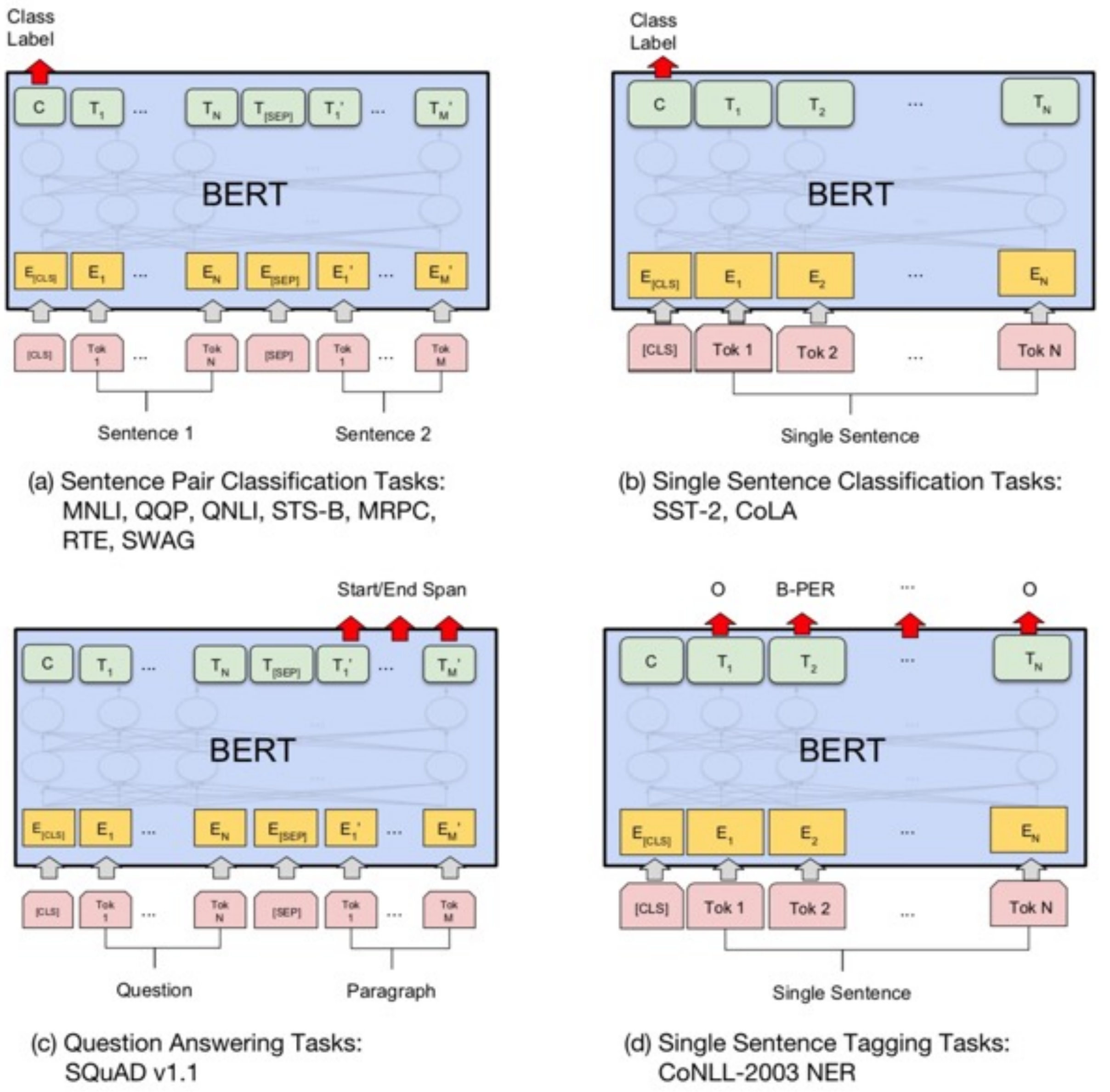
BERT (Bidirectional Encoder Representations from Transformers)

BERT input representation



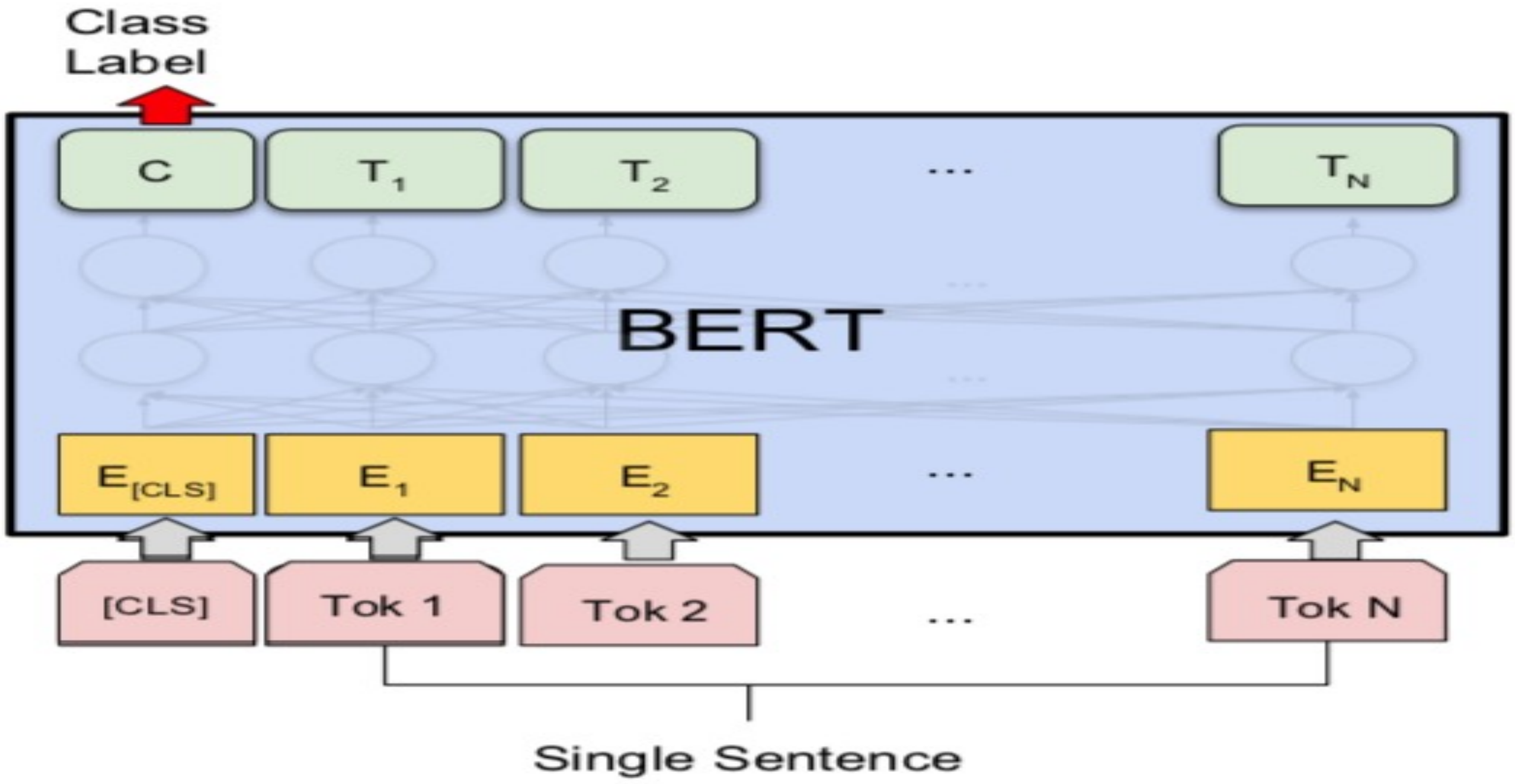
Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018).
"Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

Fine-tuning BERT on Different Tasks



Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018).
"Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

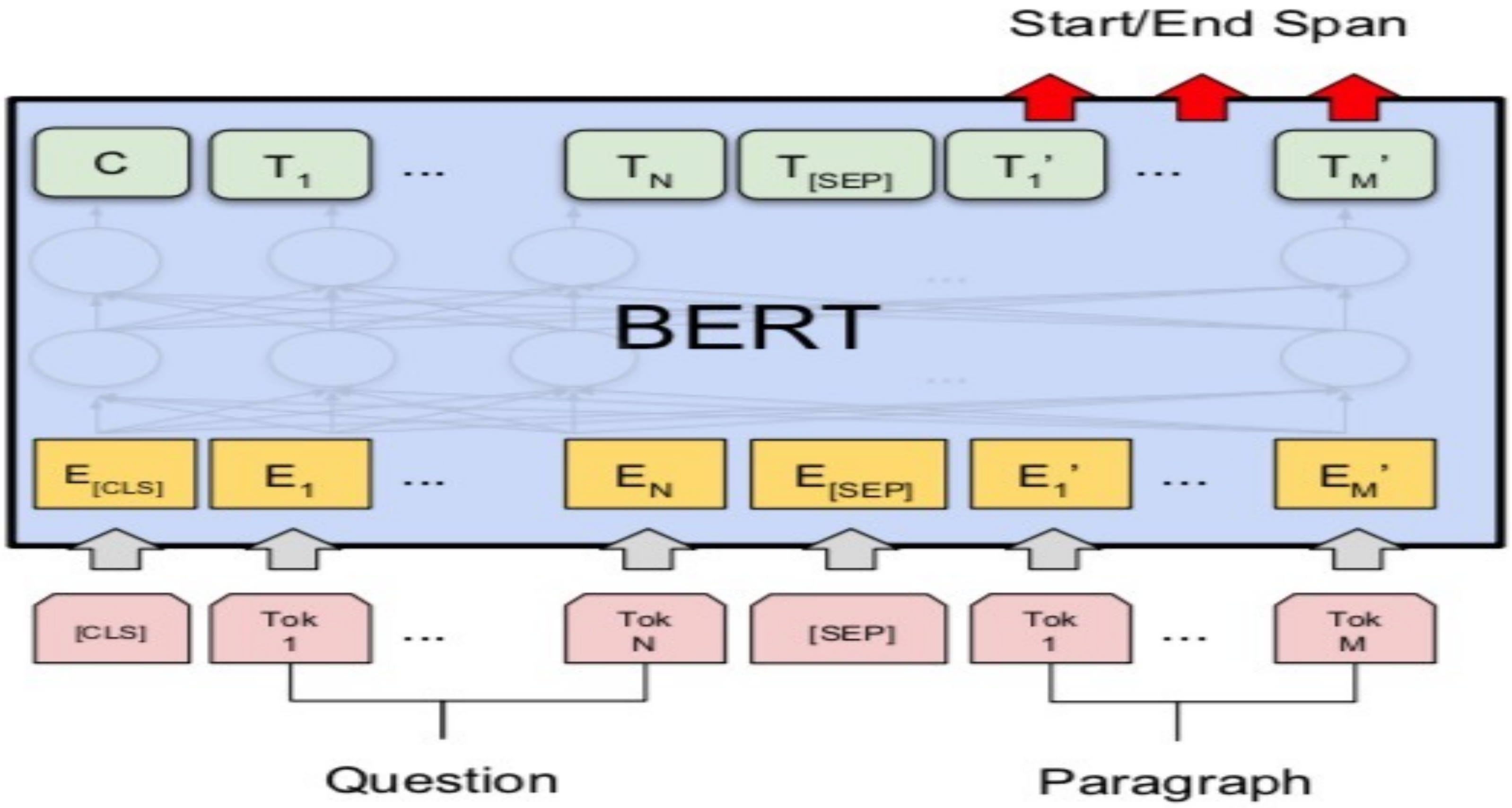
Sentiment Analysis:
Single Sentence Classification



(b) Single Sentence Classification Tasks:
SST-2, CoLA

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018).
"BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805

Fine-tuning BERT on Question Answering (QA)

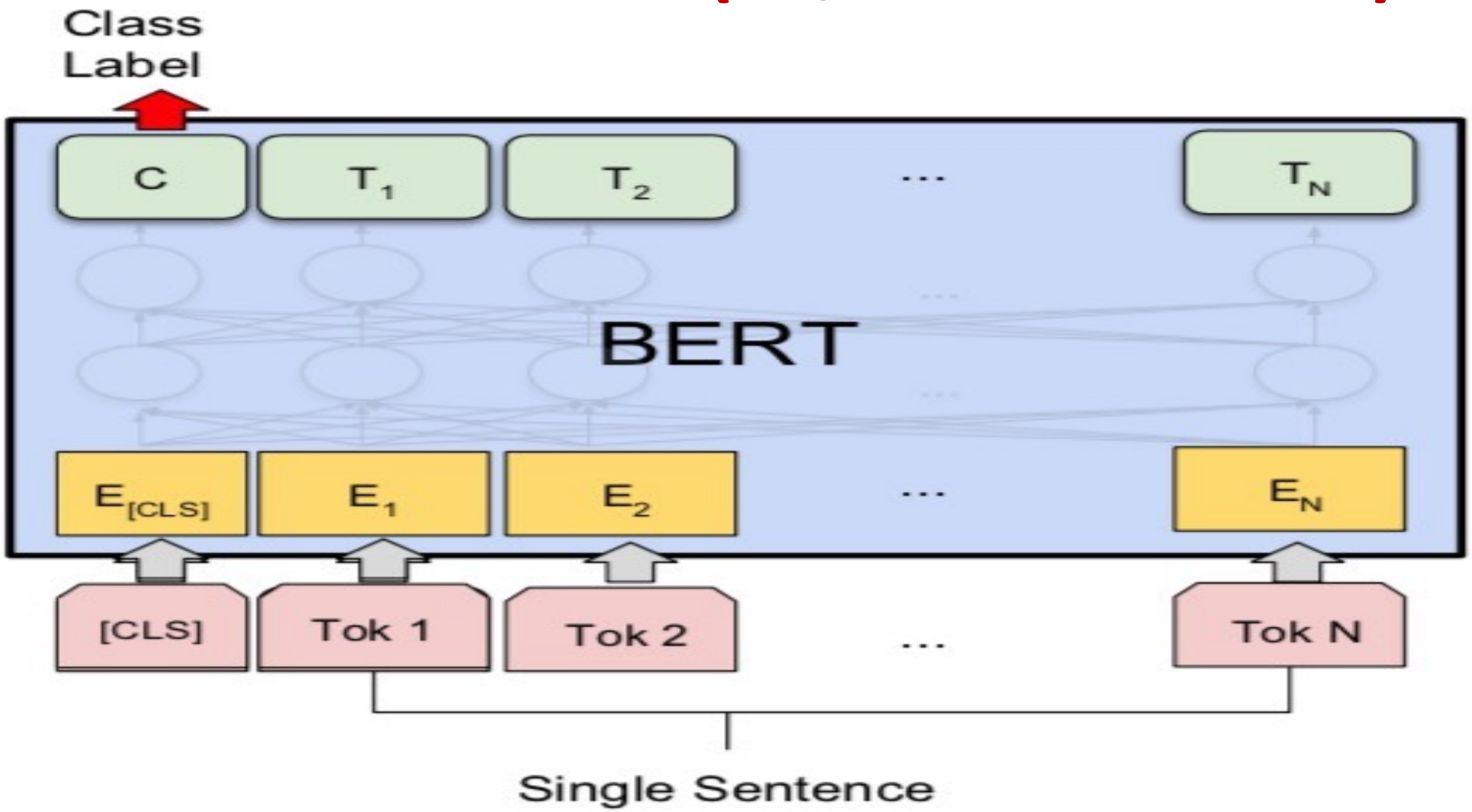


(c) Question Answering Tasks:
SQuAD v1.1

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

Fine-tuning BERT on Dialogue

Intent Detection (ID; Classification)

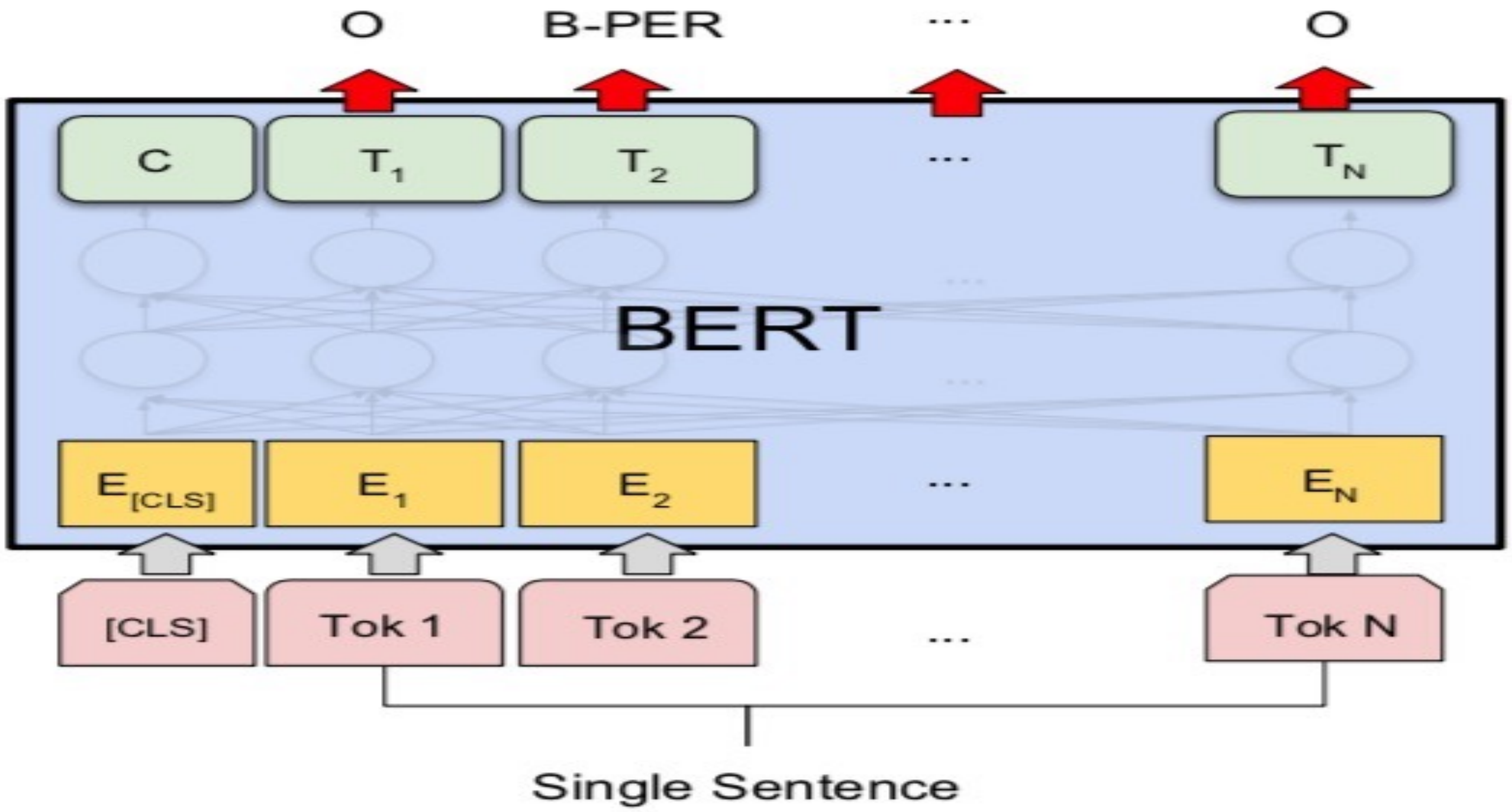


(b) Single Sentence Classification Tasks:
SST-2, CoLA

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018).
"Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

Fine-tuning BERT on Dialogue

Slot Filling (SF)

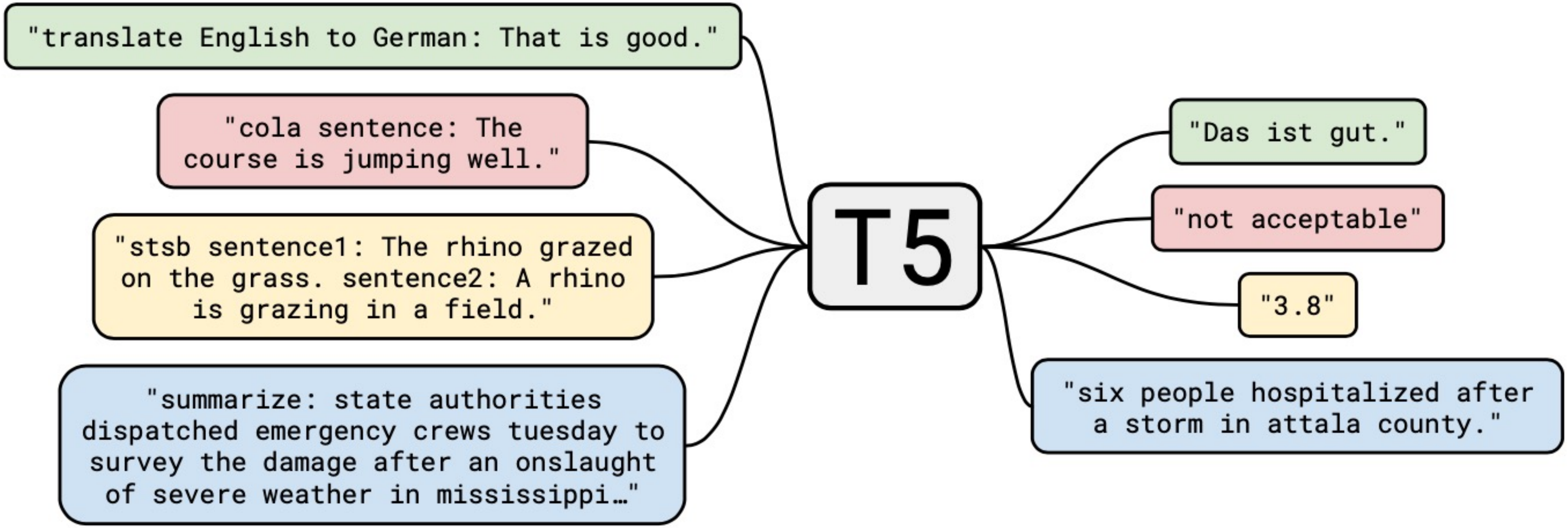


(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

Source: Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

T5

Text-to-Text Transfer Transformer



Source: JColin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu (2019). "Exploring the limits of transfer learning with a unified text-to-text transformer." arXiv preprint arXiv:1910.10683 (2019).

Hugging Face

Hugging Face | Search models, datas | Models | Datasets | Spaces | Docs | Solutions | Pricing | Log In | Sign Up

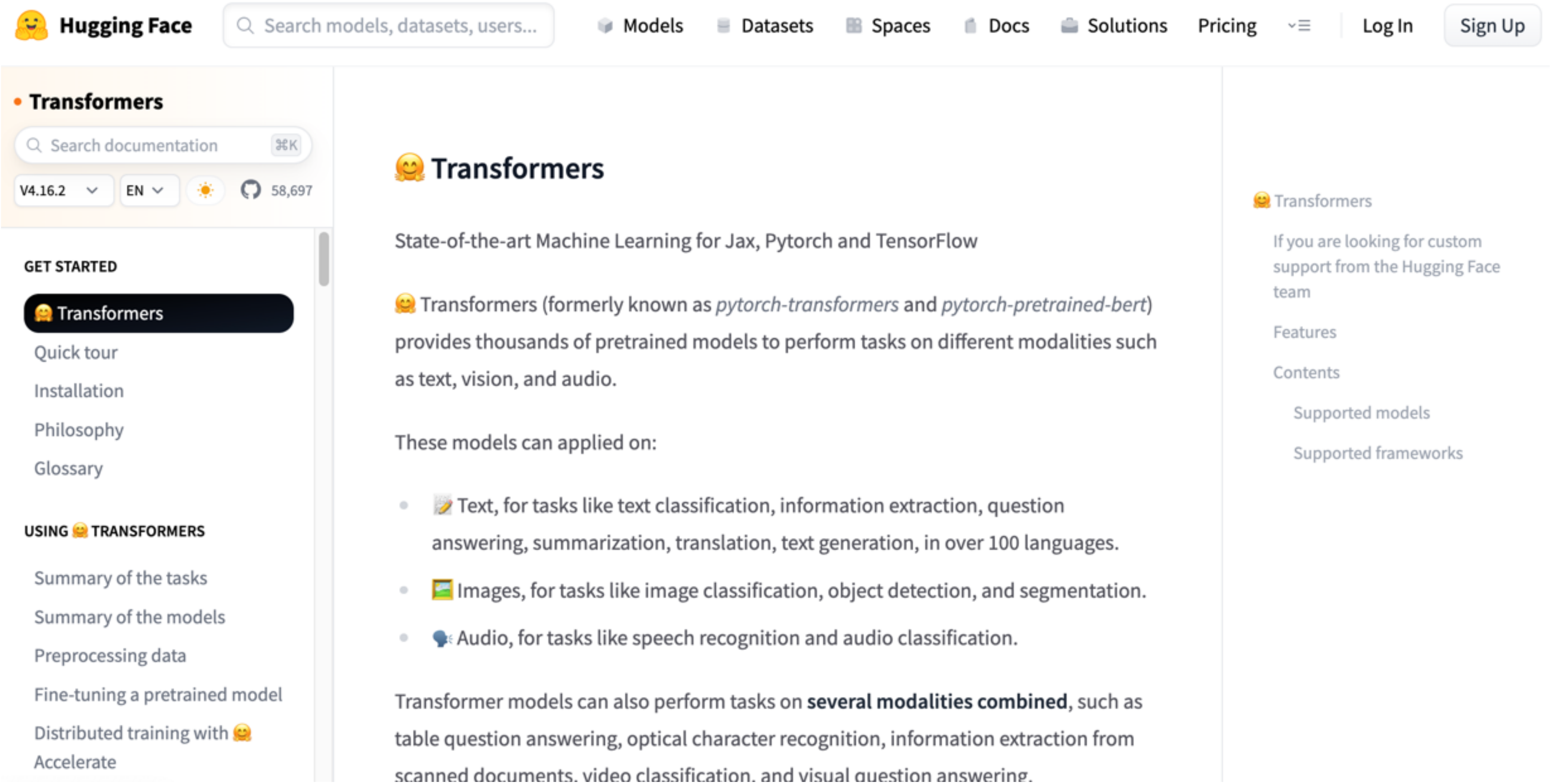
The AI community building the future.

Build, train and deploy state of the art models powered by the reference open source in machine learning.

Star 58,696

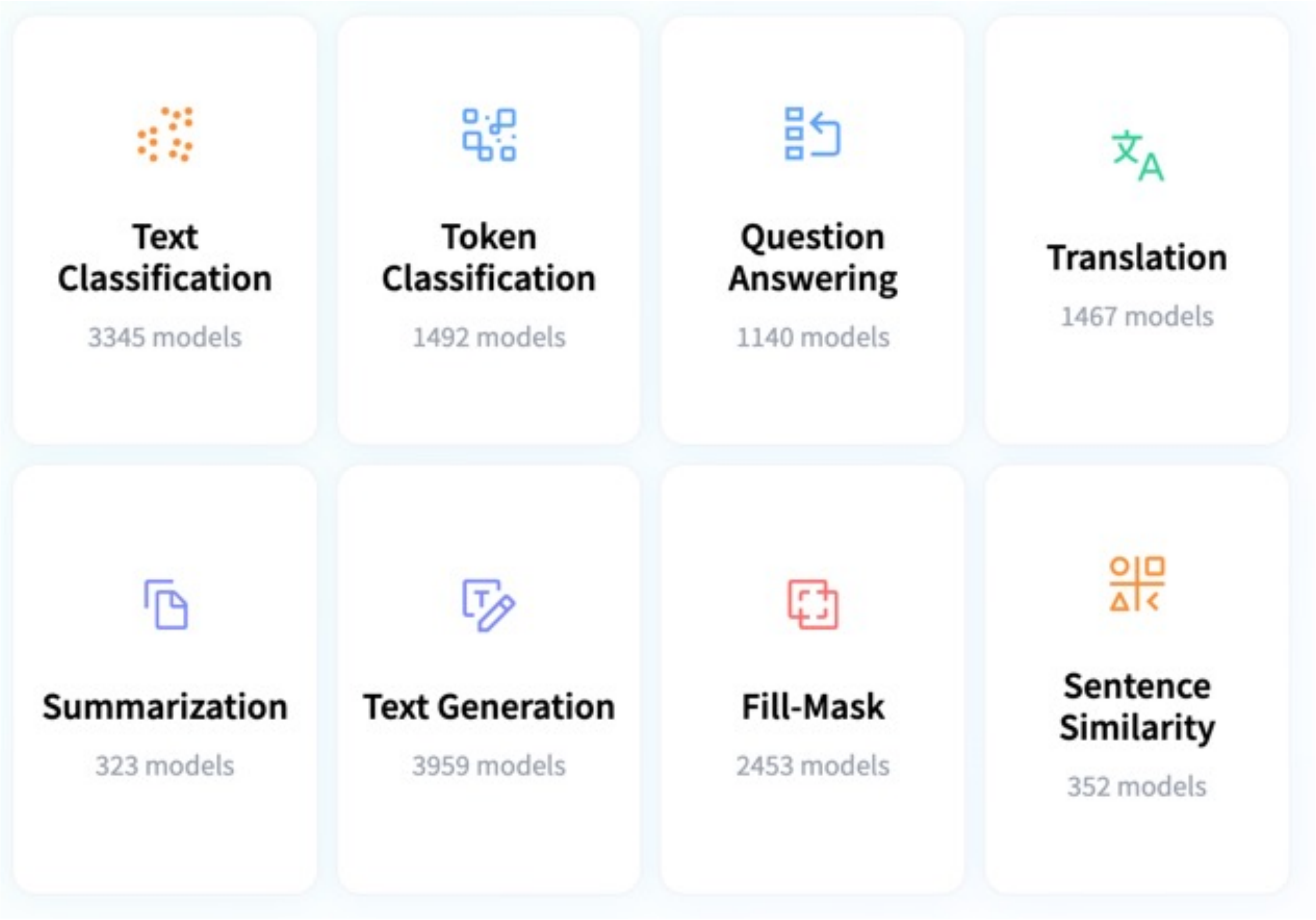
https://huggingface.co/

Hugging Face Transformers



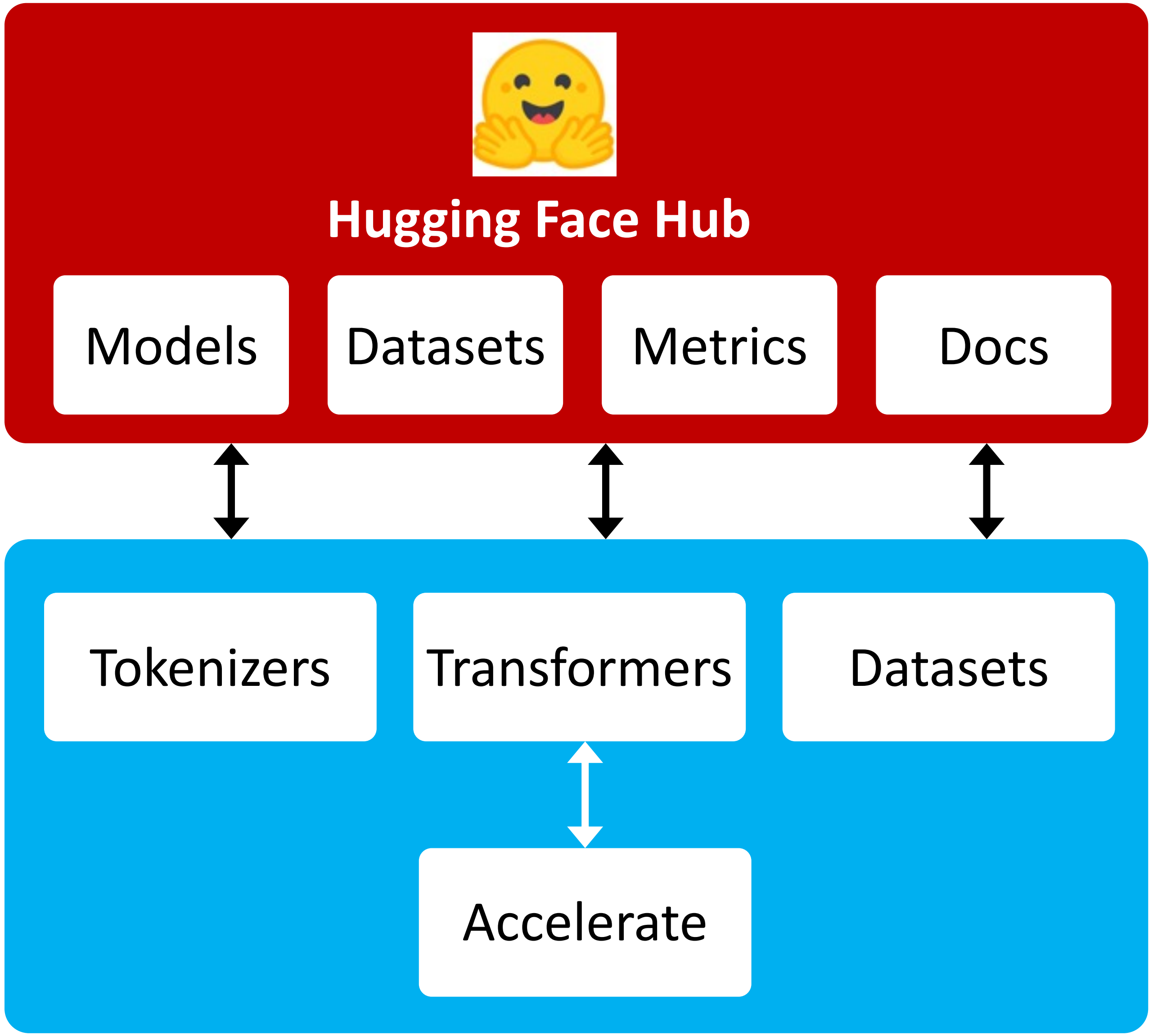
https://huggingface.co/docs/transformers/index

Hugging Face Tasks
Natural Language Processing



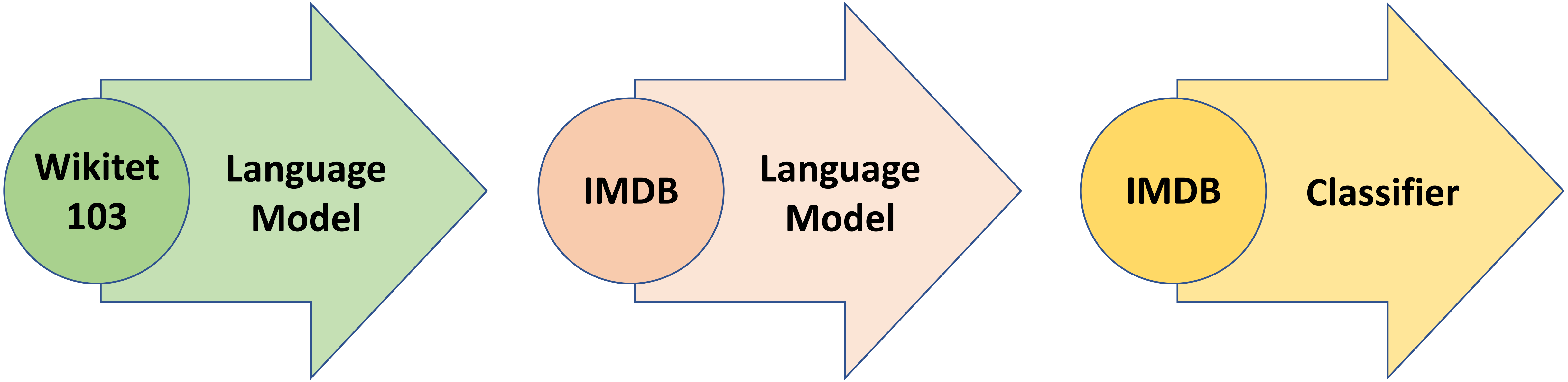
https://huggingface.co/tasks

An overview of the Hugging Face Ecosystem



Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.

ULMFiT: 3 Steps
Transfer Learning in NLP



1. Pretraining 2. Domain adaptation 3. Fine-tuning

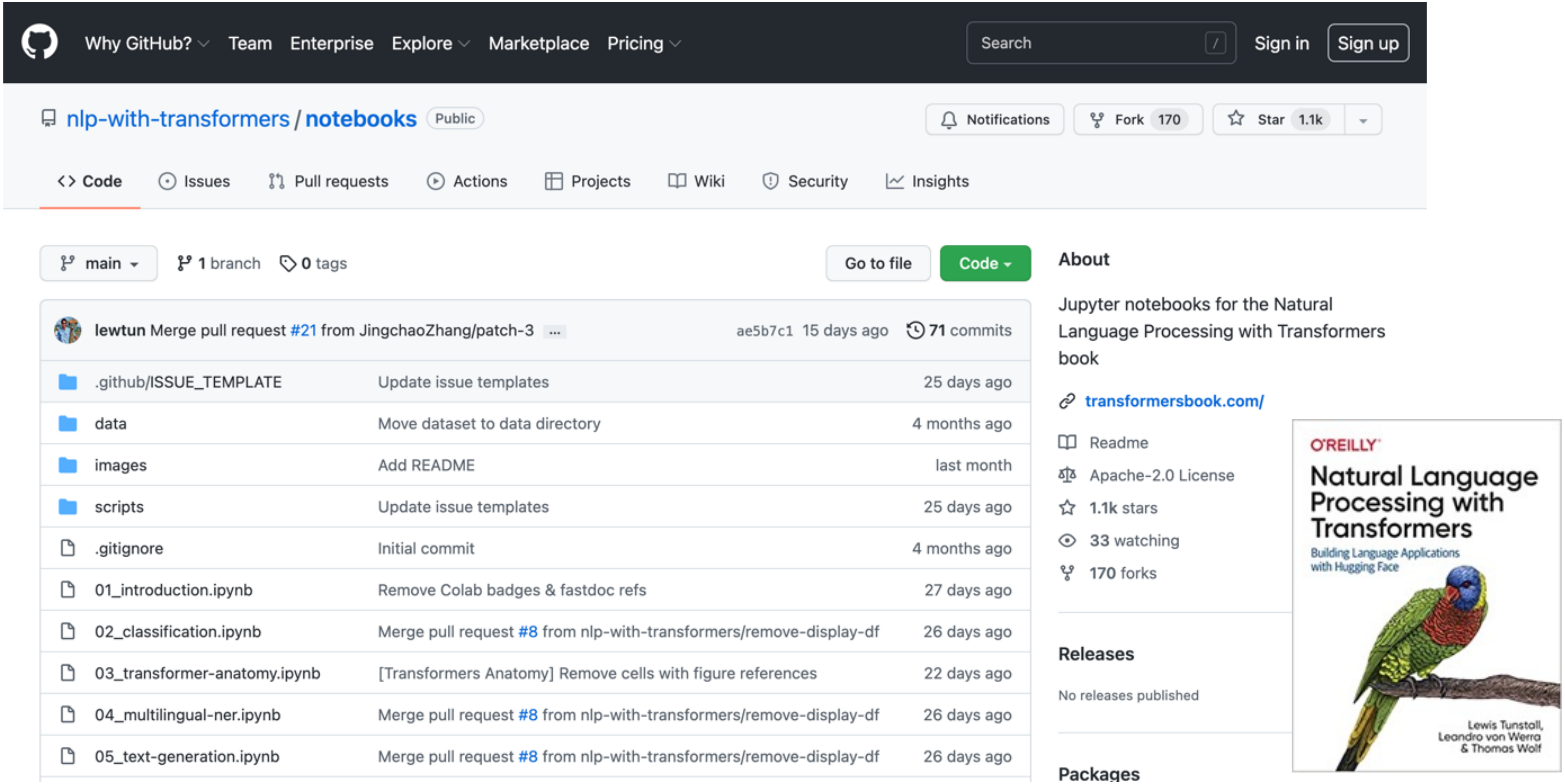
A typical pipeline for training transformer models

with the Datasets, Tokenizers, and Transformers libraries

Datasets	Tokenizers	Transformers	Datasets
Load and process datasets	Tokenize input texts	Load models, train and infer	Load metrics evaluate models

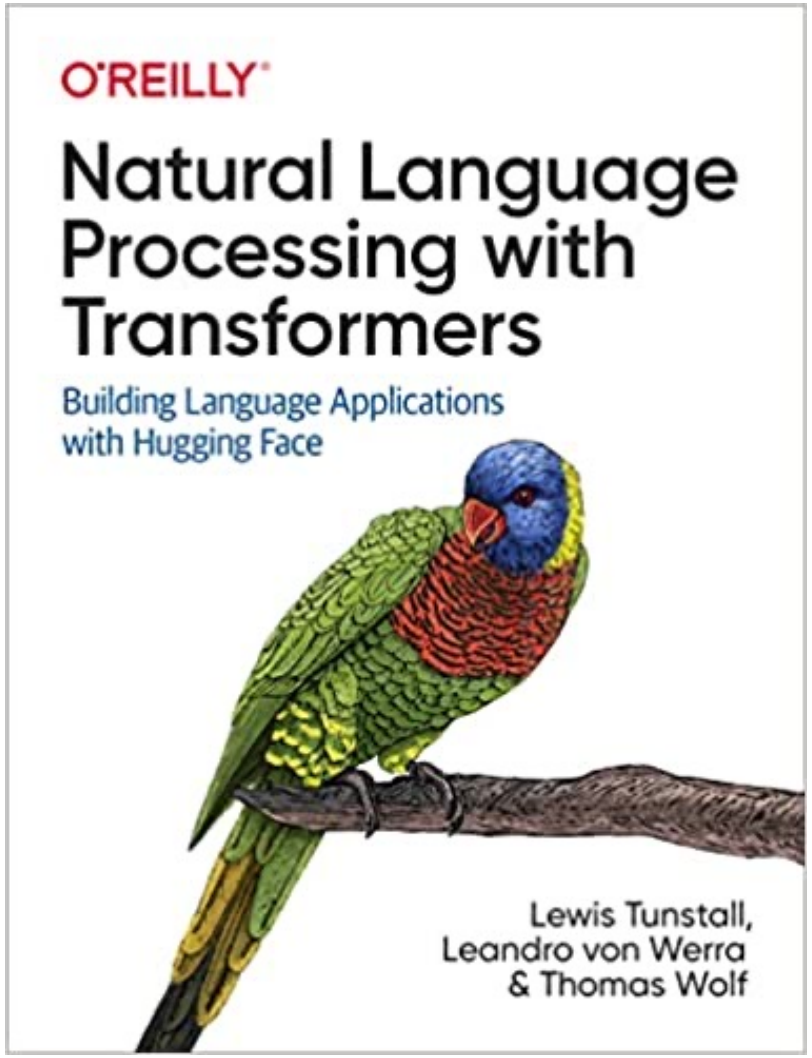
Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.
https://github.com/nlp-with-transformers/notebooks

NLP with Transformers Github



https://github.com/nlp-with-transformers/notebooks

NLP with Transformers Github Notebooks



Running on a cloud platform

To run these notebooks on a cloud platform, just click on one of the badges in the table below:

Chapter	Colab	Kaggle	Gradient	Studio Lab
Introduction	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Text Classification	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Transformer Anatomy	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Multilingual Named Entity Recognition	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Text Generation	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Summarization	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Question Answering	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Making Transformers Efficient in Production	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Dealing with Few to No Labels	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Training Transformers from Scratch	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab
Future Directions	Open in Colab	Open in Kaggle	Run on Gradient	Open Studio Lab

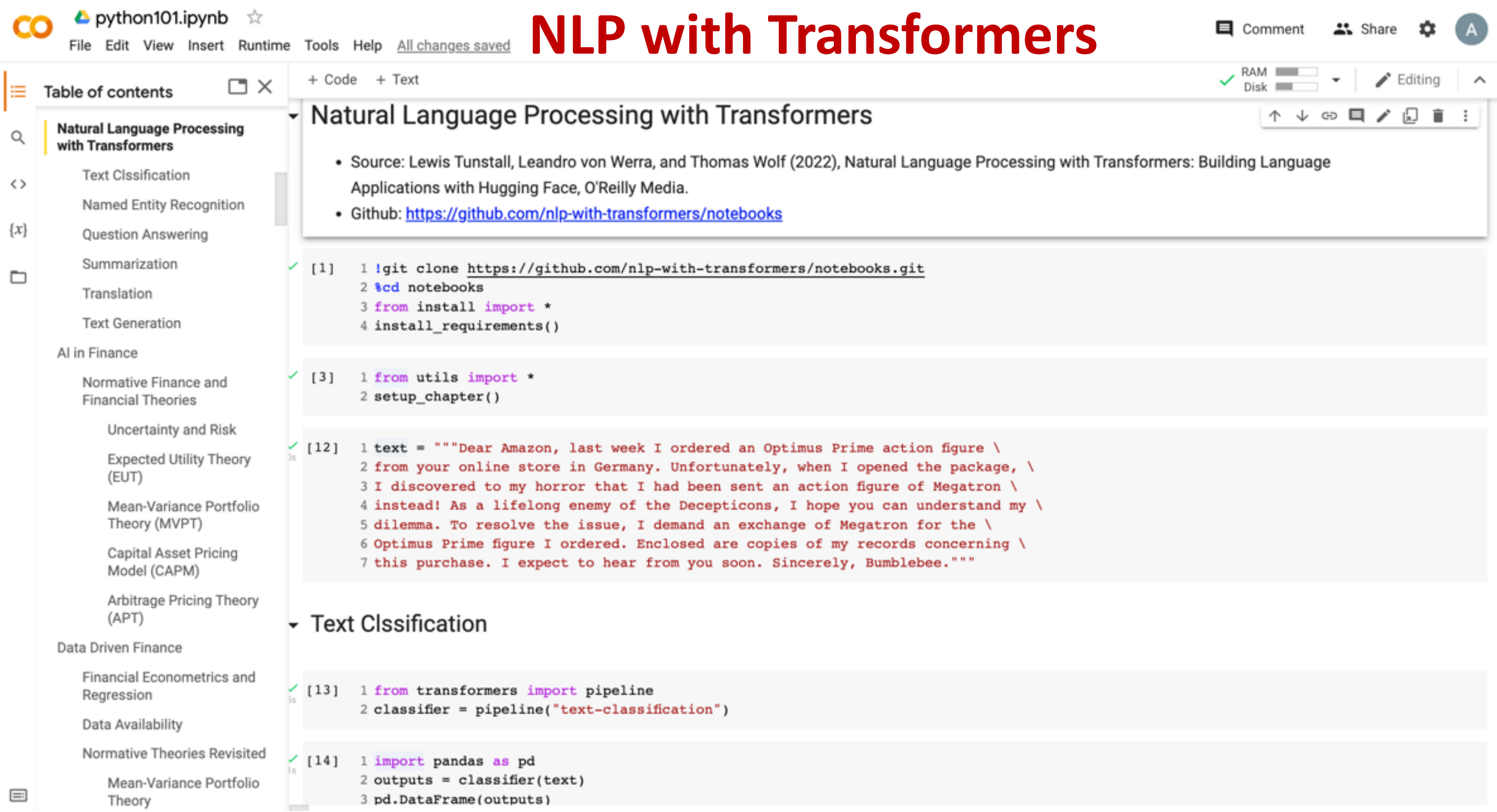
Nowadays, the GPUs on Colab tend to be K80s (which have limited memory), so we recommend using Kaggle, Gradient, or SageMaker Studio Lab. These platforms tend to provide more performant GPUs like P100s, all for free!

https://github.com/nlp-with-transformers/notebooks

Python in Google Colab (Python101)

https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT

NLP with Transformers



https://tinyurl.com/aintpupython101

Text Classification

```
!pip install transformers
from transformers import pipeline
classifier = pipeline("text-classification")
```

```
text = """Dear Amazon, last week I ordered an Optimus Prime action figure \
from your online store in Germany. Unfortunately, when I opened the package, \
I discovered to my horror that I had been sent an action figure of Megatron \
instead! As a lifelong enemy of the Decepticons, I hope you can understand my \
dilemma. To resolve the issue, I demand an exchange of Megatron for the \
Optimus Prime figure I ordered. Enclosed are copies of my records concerning \
this purchase. I expect to hear from you soon. Sincerely, Bumblebee."""
```

```
import pandas as pd
outputs = classifier(text)
pd.DataFrame(outputs)
```

	label	score
0	NEGATIVE	0.901546

Source: Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.
https://github.com/nlp-with-transformers/notebooks

Named Entity Recognition (NER)

```
from transformers import pipeline
import pandas as pd
classifier = pipeline("ner")
text = "My name is Michael and I live in Berkeley, California."
outputs = classifier(text)
pd.DataFrame(outputs)
```

	entity	score	index	word	start	end
0	I-PER	0.998874	4	Michael	11	18
1	I-LOC	0.997050	9	Berkeley	33	41
2	I-LOC	0.999170	11	California	43	53

https://tinyurl.com/aintpupython101

Text Generation

```
!pip install transformers
from transformers import pipeline
generator = pipeline('text-generation', model = 'gpt2')
generator("Hello, I'm a language model", max_length = 30, num_return_sequences=3)
```

```
[{'generated_text': "Hello, I'm a language model. It's like looking at it,
where is each word of the sentence? That's what I mean. Like"},
{'generated_text': "Hello, I'm a language modeler. I'm using this for two
purposes: I'm having a lot fewer bugs and faster performance. If I"},
{'generated_text': 'Hello, I\'m a language model, and I was born to
code."\n\nNow, I am thinking about this from a different perspective with
a'}]
```

https://tinyurl.com/aintpupython101

Text Generation

```
from transformers import pipeline
generator = pipeline('text-generation', model = 'gpt2')
outputs = generator("Once upon a time", max_length = 30)
print(outputs[0]['generated_text'])
```

```
Once upon a time, every person who ever saw
Jesus, knew that He was Christ. And even
though he might not have known Him, He was
```

Text Generation

```
from transformers import pipeline
generator = pipeline('text-generation', model = 'gpt2')
outputs = generator("Once upon a time", max_length = 100)
print(outputs[0]['generated_text'])
```

```
Once upon a time we should be able to speak to people who
have lost children, so we try to take those that have lost
the children to our institutions — but the first time is very
hard for us because of our institutions. To me, it's
important to acknowledge that in an institution of faith and
love they are not children. And that there are many people
who are still hurting the child and there are many in need of
help, if not a system. So I'm very curious
```

Text2Text Generation

```
from transformers import pipeline
text2text_generator = pipeline("text2text-generation", model = 't5-base')
outputs = text2text_generator("translate from English to French: I am a student")
print(outputs[0]['generated_text'])
```

I am a student

Je suis un étudiant

https://tinyurl.com/aintpupython101

Text2Text Generation

```python
from transformers import pipeline
text2text_generator = pipeline("text2text-generation")
text2text_generator("question: What is 42 ? context: 42 is the answer to life, the universe and everything")
```

```
[{'generated_text': 'the answer to life, the universe and everything'}]
```

https://tinyurl.com/aintpupython101

Question Answering

```
!pip install transformers
from transformers import pipeline
qamodel = pipeline("question-answering")
question = "Where do I live?"
context = "My name is Michael and I live in Taipei."
qamodel(question = question, context = context)
```

```
{'answer': 'Taipei', 'end': 39, 'score': 0.9730741381645203, 'start': 33}
```

https://tinyurl.com/aintpupython101

Question Answering

```python
from transformers import pipeline
qamodel = pipeline("question-answering", model ='deepset/roberta-base-squad2')
question = "Where do I live?"
context = "My name is Michael and I live in Taipei."
output = qamodel(question = question, context = context)
print(output['answer’])
```

```
Taipei
```

https://tinyurl.com/aintpupython101

Question Answering

```
from transformers import pipeline
qamodel = pipeline("question-answering", model ='deepset/roberta-base-squad2')
question = "What causes precipitation to fall?"
context = """In meteorology, precipitation is any product of
the condensation of atmospheric water vapor that falls under
gravity. The main forms of precipitation include drizzle,
rain, sleet, snow, graupel and hail... Precipitation forms as
smaller droplets coalesce via collision with other rain drops
or ice crystals within a cloud. Short, intense periods of
rain in scattered locations are called "showers"."""
output = qamodel(question = question, context = context)
print(output['answer'])
```

```
gravity
```

https://tinyurl.com/aintpupython101

Outline

- AI for Text Analytics
  - Natural Language Processing with Transformers: Building Language Applications with Hugging Face
  - Practical Natural Language Processing
- FinTech: Financial Services Innovation
- Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech

FinTech: Financial Services Innovation

FinTech

Financial Technology
FinTech

“providing
financial services
by making use of
software and
modern technology”

Source: https://www.fintechweekly.com/fintech-definition

Financial Services

Financial Services

Source: http://www.crackitt.com/7-reasons-why-your-fintech-startup-needs-visual-marketing/

Paolo Sironi (2016)

FinTech Innovation:

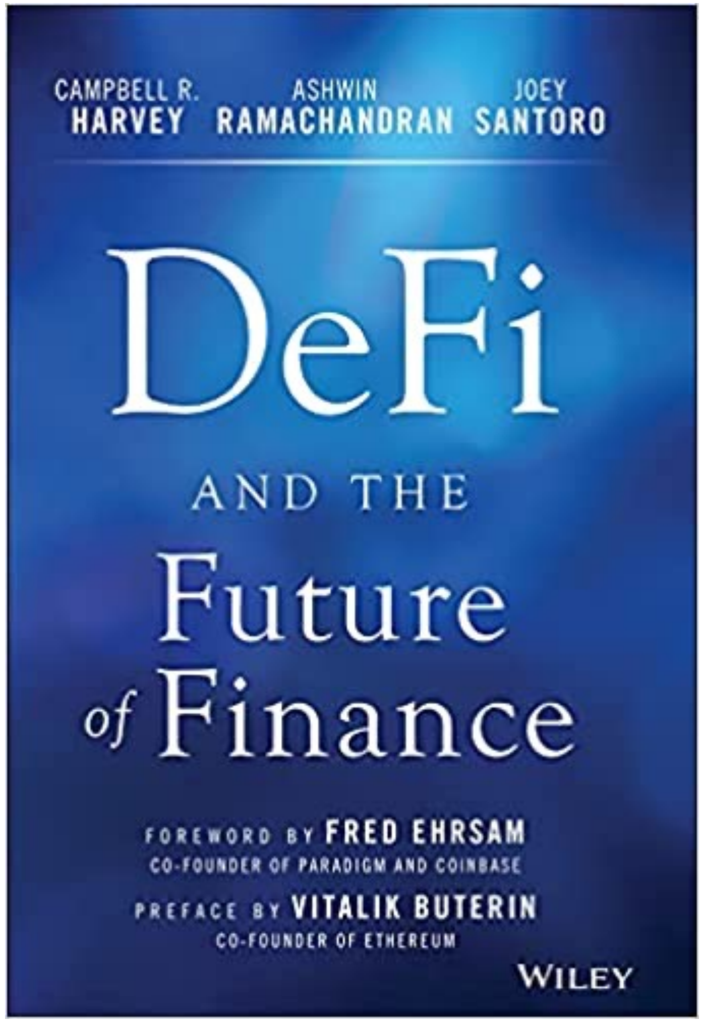
From Robo-Advisors to Goal Based Investing and Gamification, Wiley

Source: https://www.amazon.com/FinTech-Innovation-Robo-Advisors-Investing-Gamification/dp/1119226988

Campbell R. Harvey, Ashwin Ramachandran, Joey Santoro, Fred Ehrsam (2021),

DeFi and the Future of Finance,

Wiley



Source: https://www.amazon.com/DeFi-Future-Finance-Campbell-Harvey-ebook/dp/B09DJV2QLC

FinTech: Financial Services Innovation

Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf

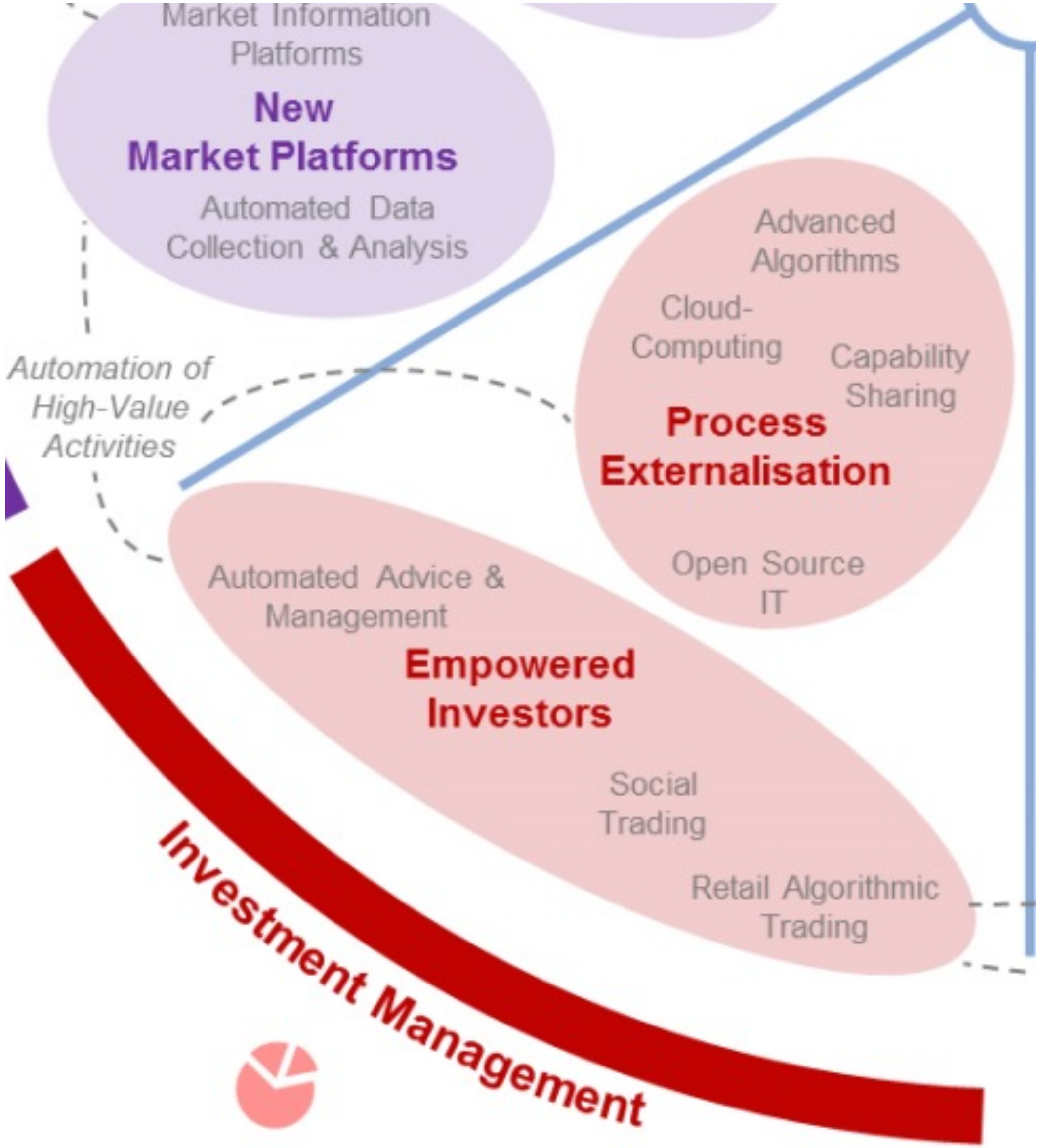
FinTech:

Financial Services Innovation

1. Payments
2. Insurance
3. Deposits & Lending
4. Capital Raising
5. Investment Management
6. Market Provisioning

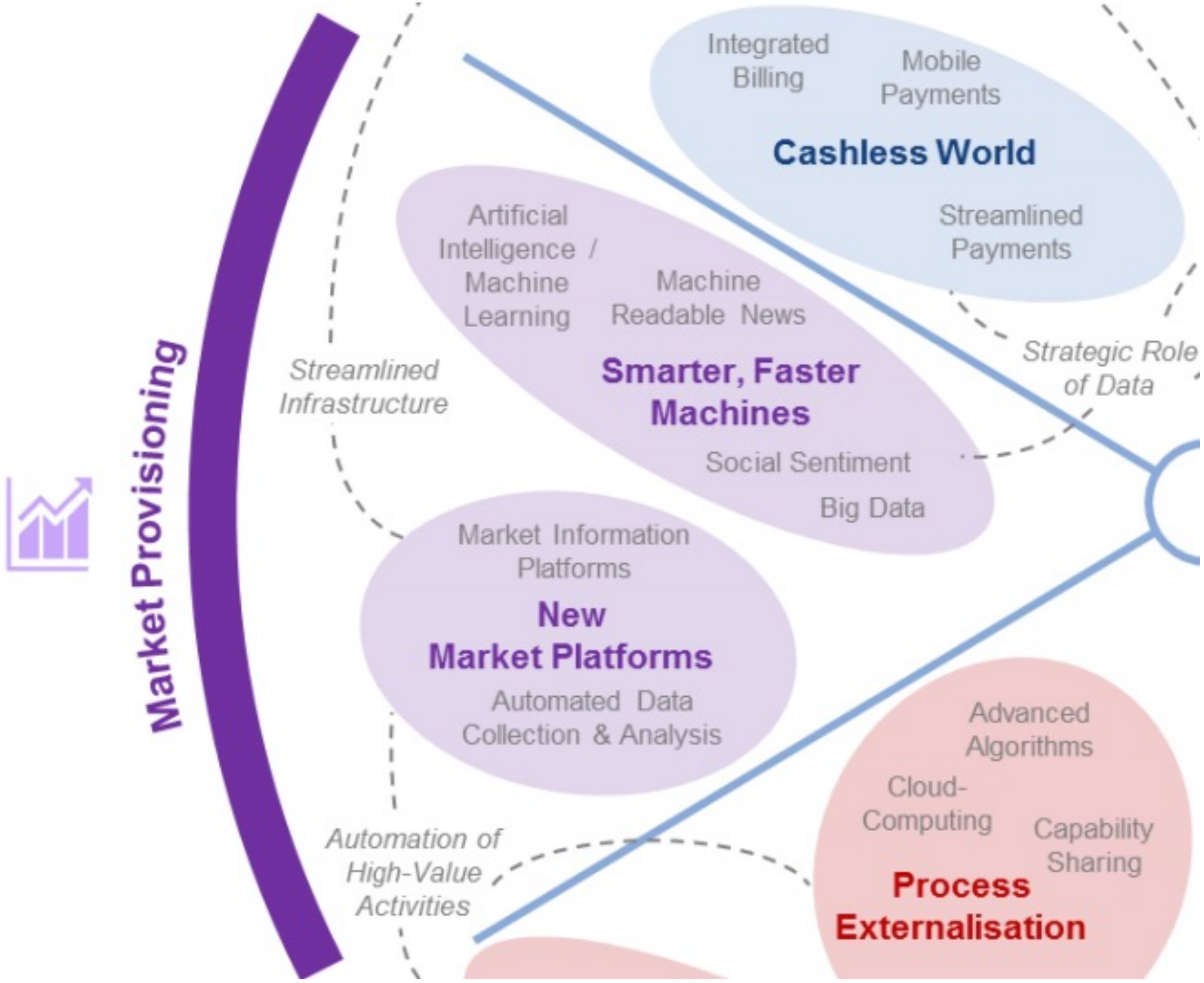
Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf

5 FinTech: Investment Management



Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf

6 FinTech: Market Provisioning



Source: http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf

FinTech ABCD

AI

Block Chain

Cloud Computing

Big **D**ata

Decentralized Finance (DeFi)

Block Chain Financial Technology

Block Chain & Bitcoin (BTC)

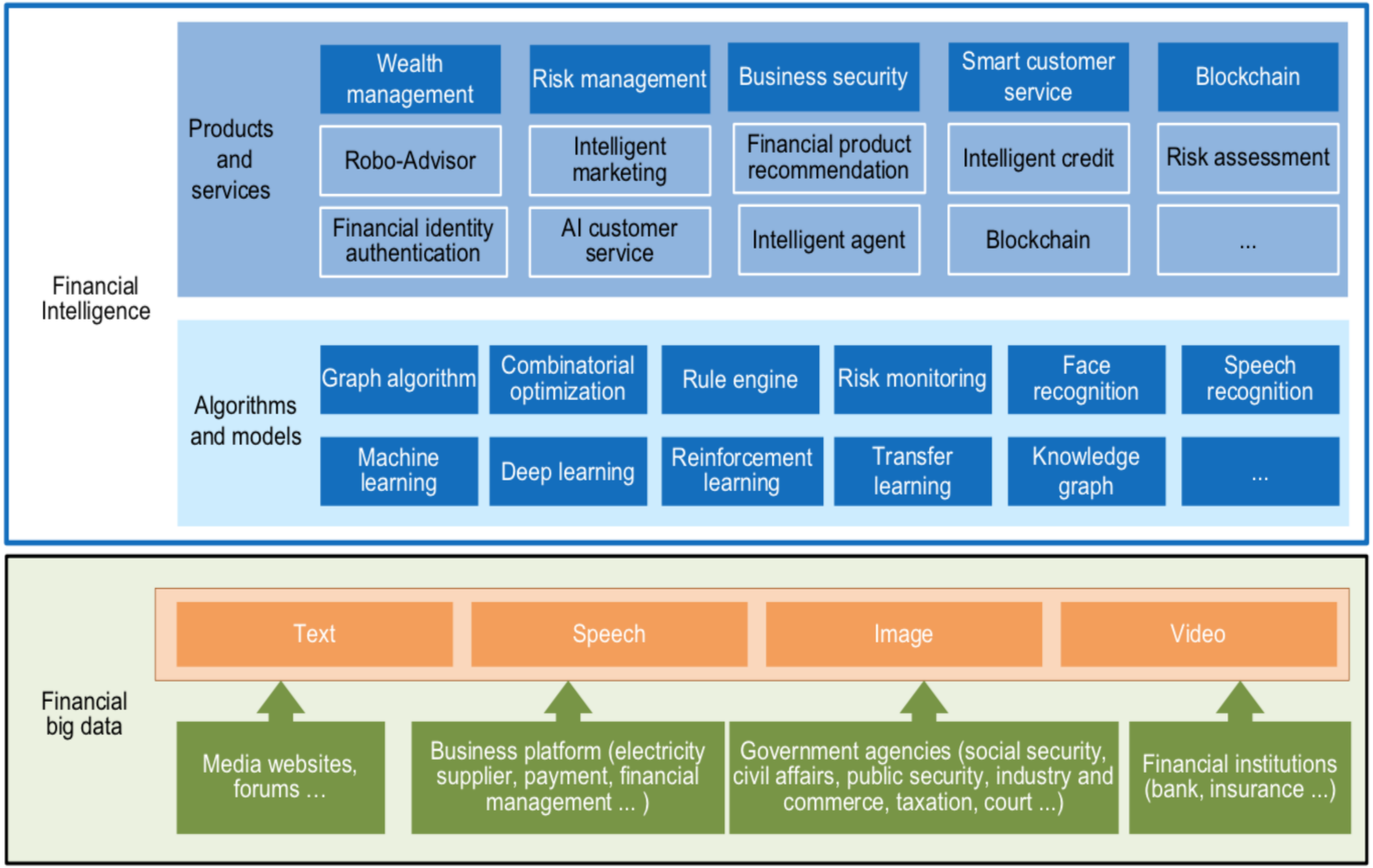
Smart Contract & Ethereum (ETH)

Decentralized Application (DApp)

AI
in
FinTech

FinBrain: when Finance meets AI 2.0

(Zheng et al., 2019)



Source: Xiao-lin Zheng, Meng-ying Zhu, Qi-bing Li, Chao-chao Chen, and Yan-chao Tan (2019), "Finbrain: When finance meets AI 2.0." Frontiers of Information Technology & Electronic Engineering 20, no. 7, pp. 914-924

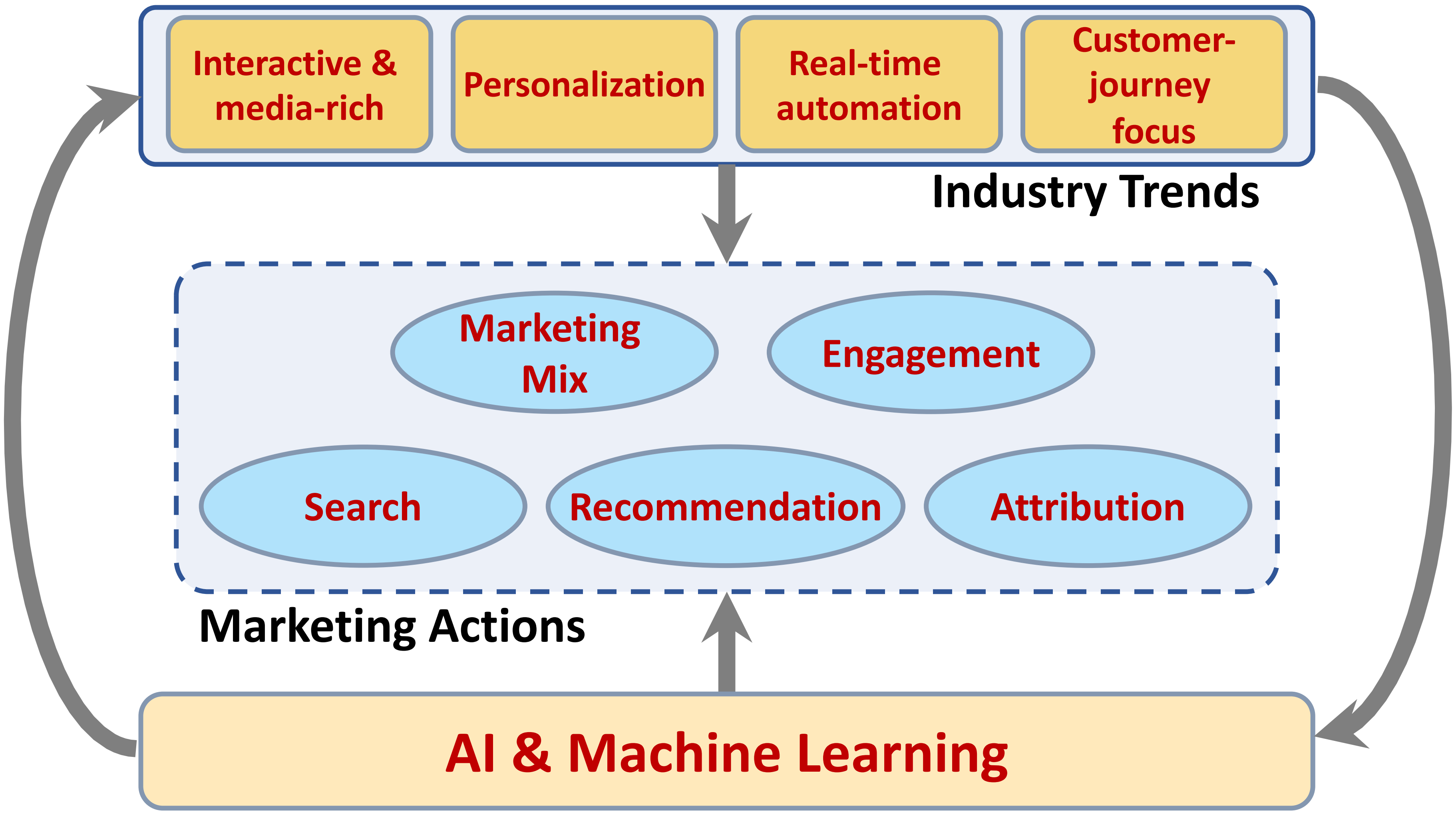
Technology-driven Financial Industry Development

Development stage	Driving technology	Main landscape	Inclusive finance	Relationship between technology and finance
Fintech 1.0 (financial IT)	Computer	Credit card, ATM, and CRMS	Low	Technology as a tool
Fintech 2.0 (Internet finance)	Mobile Internet	Marketplace lending, third-party payment, crowdfunding, and Internet insurance	Medium	Technology-driven change
Fintech 3.0 (financial intelligence)	AI, Big Data, Cloud Computing, Blockchain	Intelligent finance	High	Deep fusion

Source: Xiao-lin Zheng, Meng-ying Zhu, Qi-bing Li, Chao-chao Chen, and Yan-chao Tan (2019), "Finbrain: When finance meets AI 2.0." Frontiers of Information Technology & Electronic Engineering 20, no. 7, pp. 914-924

AI-driven Marketing

(Ma and Sun, 2020)



Source: Liye Ma and Baohong Sun (2020), "Machine learning and AI in marketing – Connecting computing power to human insights." International Journal of Research in Marketing, 37, no. 3, 481-504.

Deep learning for financial applications: A survey

Applied Soft Computing (2020)

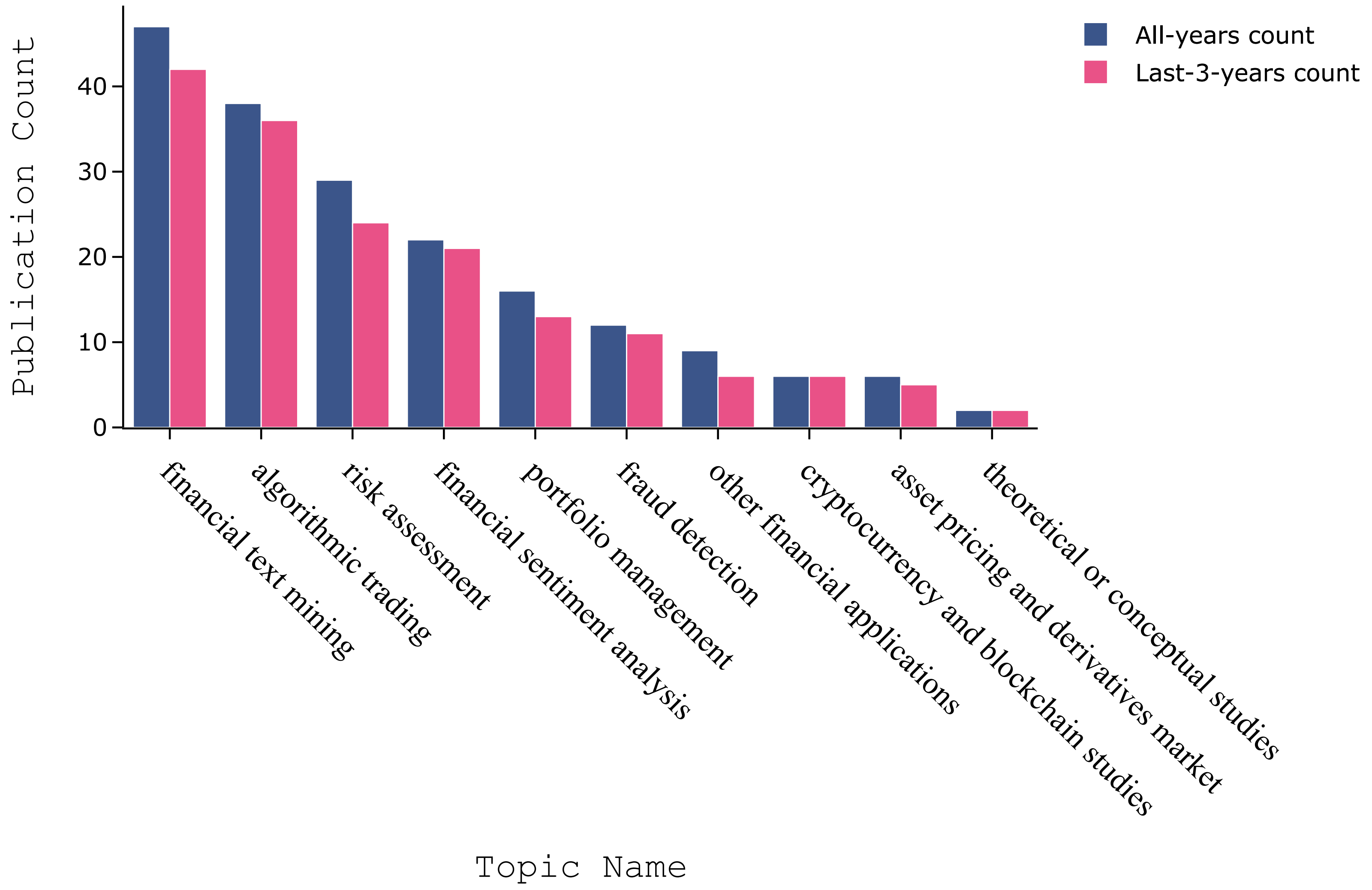
Source:
Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey."
Applied Soft Computing (2020): 106384.

Financial time series forecasting with deep learning: A systematic literature review: 2005–2019

Applied Soft Computing (2020)

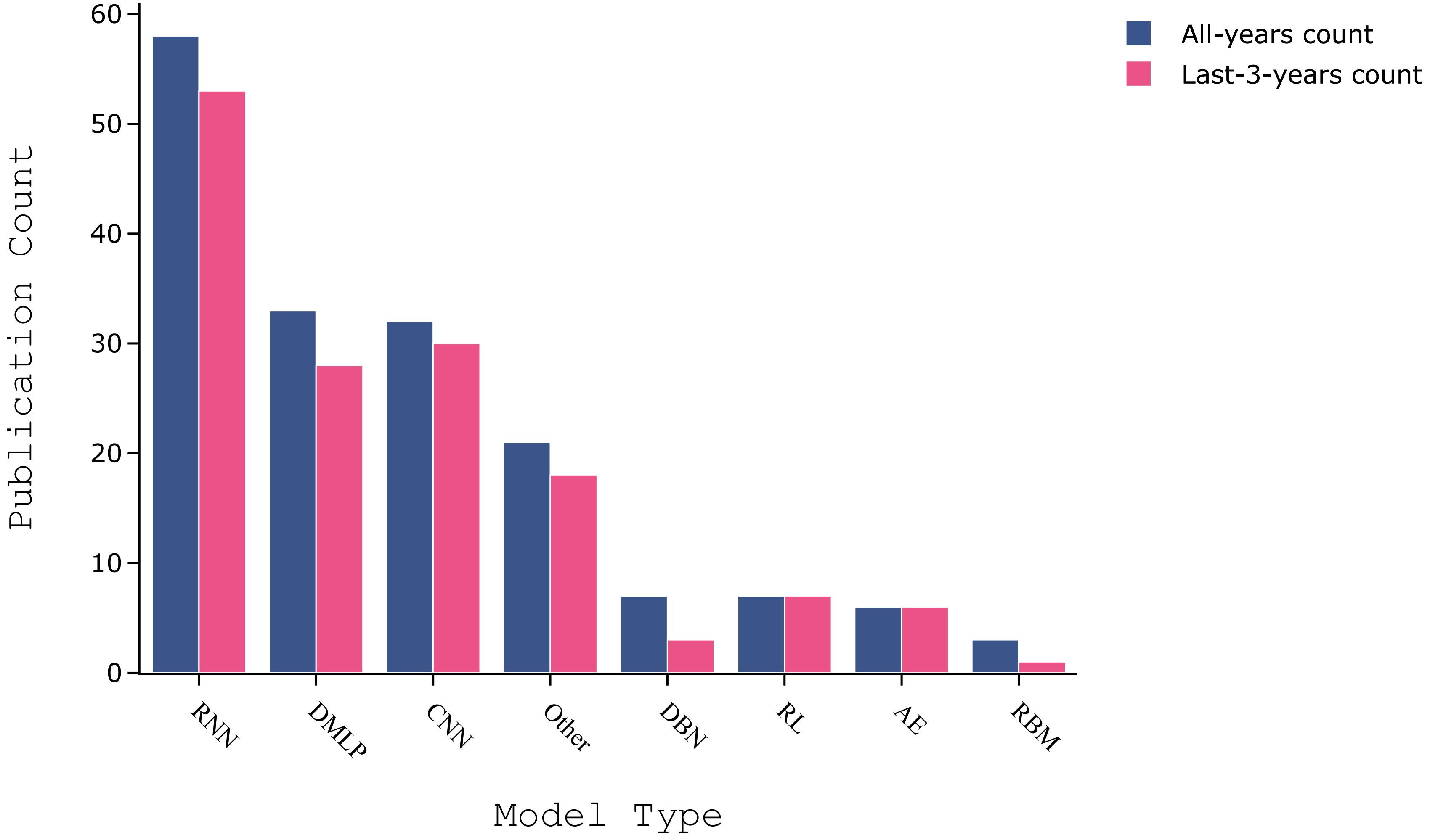
Source:
Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu (2020), "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019." Applied Soft Computing 90 (2020): 106181.

Deep learning for financial applications: Topics



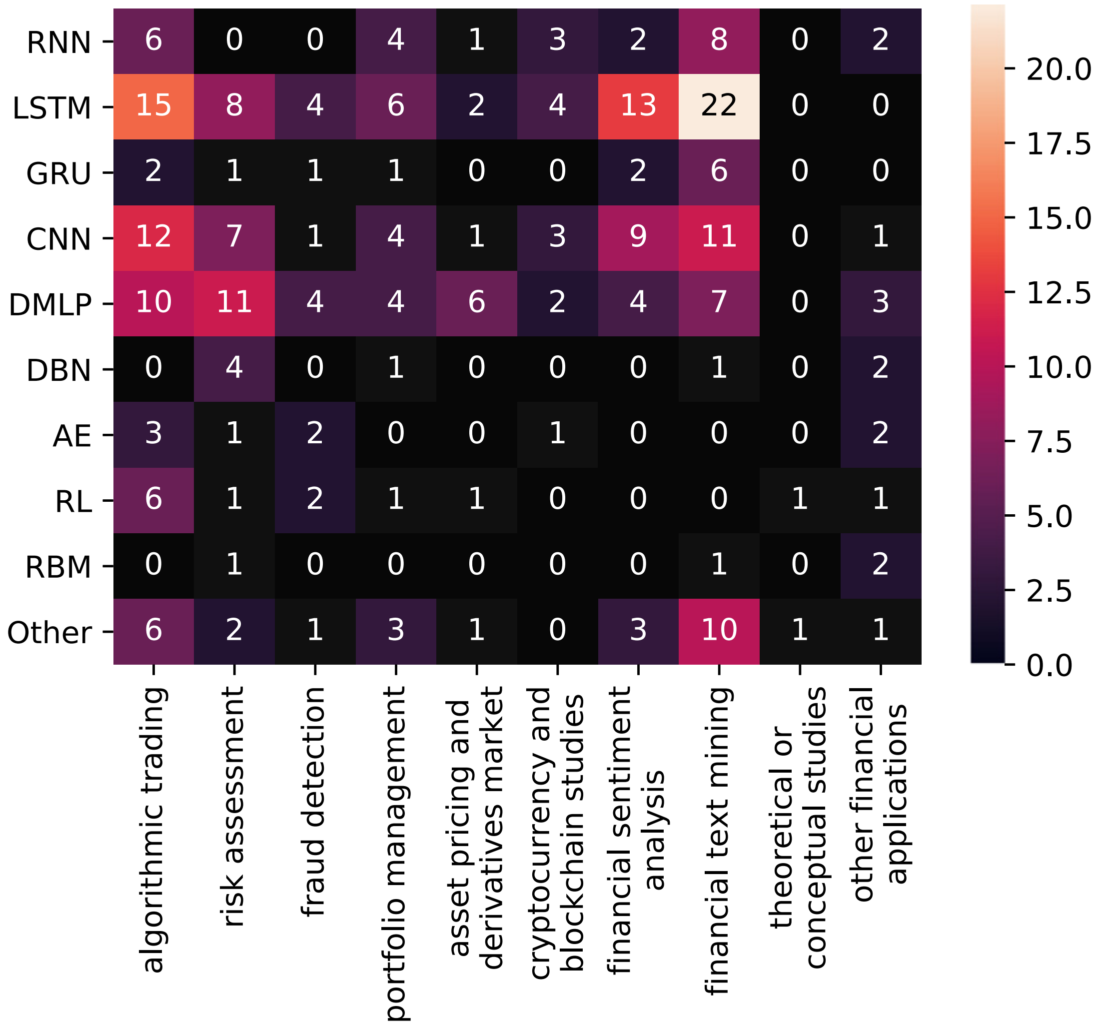
Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications: Deep Learning Models



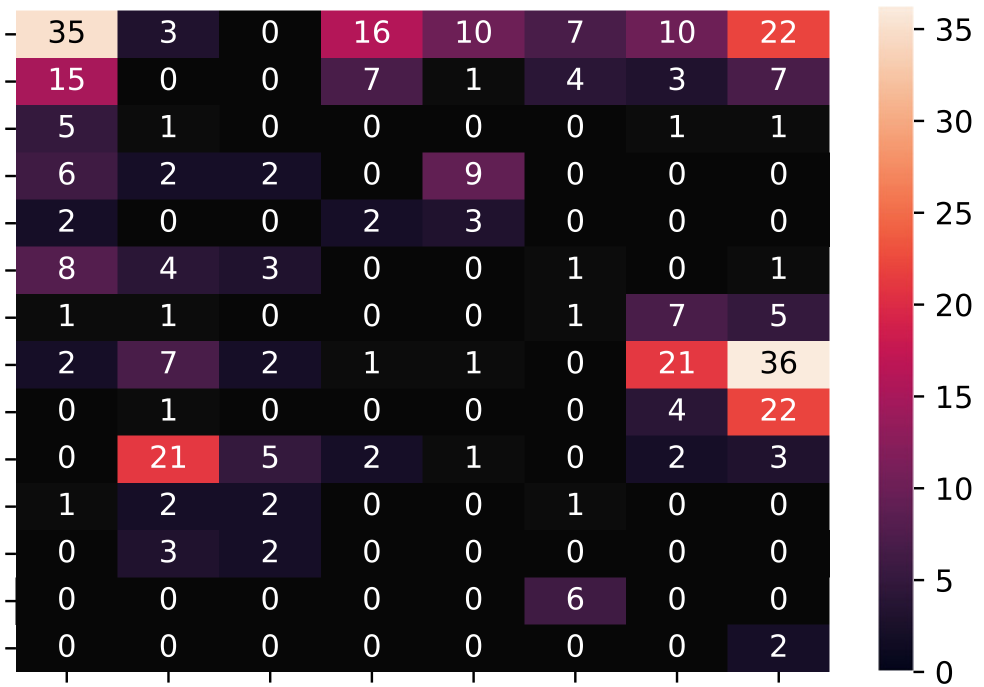
Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications: Topic-Model Heatmap



Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

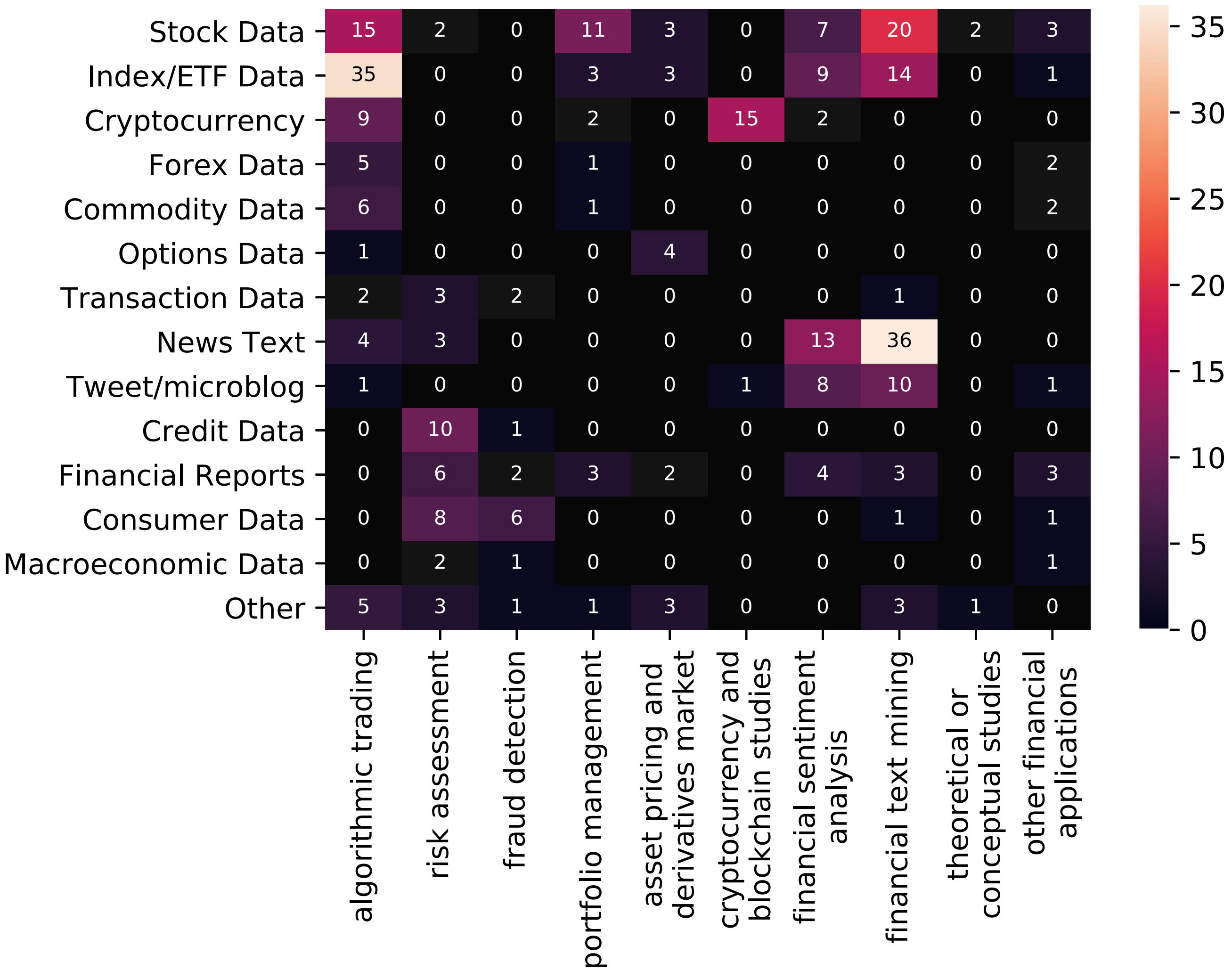
Deep learning for financial applications: Topic-Feature Heatmap



	algorithmic trading	risk assessment	fraud detection	portfolio management	asset pricing and derivatives market	cryptocurrency and blockchain studies	financial sentiment analysis	financial text mining
price data	35	3	0	16	10	7	10	22
technical indicator	15	0	0	7	1	4	3	7
index data	5	1	0	0	0	0	1	1
market characteristics	6	2	2	0	9	0	0	0
fundamental	2	0	0	2	3	0	0	0
market microstructure data	8	4	3	0	0	1	0	1
sentiment	1	1	0	0	0	1	7	5
text	2	7	2	1	1	0	21	36
news	0	1	0	0	0	0	4	22
company/personal financial data	0	21	5	2	1	0	2	3
macroeconomic data	1	2	2	0	0	1	0	0
risk measuring features	0	3	2	0	0	0	0	0
blockchain/cryptocurrency specific features	0	0	0	0	0	6	0	0
human inputs	0	0	0	0	0	0	0	2

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications: Topic-Dataset Heatmap



Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Financial sentiment studies coupled with text mining for forecasting

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[137]	Analyst reports on the TSE and Osaka Exchange	2016–2018	Text	LSTM, CNN, Bi-LSTM	Accuracy, R^2	R, Python, MeCab
[150]	Sina Weibo, Stock market records	2012–2015	Technical indicators, sentences	DRSE	F1-score, precision, recall, accuracy, AUROC	Python
[151]	News from Reuters and Bloomberg for S&P500 stocks	2006–2015	Financial news, price data	DeepClue	Accuracy	Dynet software
[152]	News from Reuters and Bloomberg, Historical stock security data	2006–2013	News, price data	DMLP	Accuracy	–
[153]	SCI prices	2008–2015	OCHL of change rate, price	Emotional Analysis + LSTM	MSE	–
[154]	SCI prices	2013–2016	Text data and Price data	LSTM	Accuracy, F1-Measure	Python, Keras
[155]	Stocks of Google, Microsoft and Apple	2016–2017	Twitter sentiment and stock prices	RNN	–	Spark, Flume,Twitter API,
[156]	30 DJIA stocks, S&P500, DJI, news from Reuters	2002–2016	Price data and features from news articles	LSTM, NN, CNN and word2vec	Accuracy	VADER
[157]	Stocks of CSI300 index, OCHLV of CSI300 index	2009–2014	Sentiment Posts, Price data	Naive Bayes + LSTM	Precision, Recall, F1-score, Accuracy	Python, Keras
[158]	S&P500, NYSE Composite, DJIA, NASDAQ Composite	2009–2011	Twitter moods, index data	DNN, CNN	Error rate	Keras, Theano

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Text mining studies without sentiment analysis for forecasting

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[68]	Energy-Sector/ Company-Centric Tweets in S&P500	2015–2016	Text and Price data	RNN, KNN, SVR, LinR	Return, SR, precision, recall, accuracy	Python, Tweepy API
[165]	News from Reuters, Bloomberg	2006–2013	Financial news, price data	Bi-GRU	Accuracy	Python, Keras
[166]	News from Sina.com, ACE2005 Chinese corpus	2012–2016	A set of news text	Their unique algorithm	Precision, Recall, F1-score	–
[167]	CDAX stock market data	2010–2013	Financial news, stock market data	LSTM	MSE, RMSE, MAE, Accuracy, AUC	TensorFlow, Theano, Python, Scikit-Learn
[168]	Apple, Airbus, Amazon news from Reuters, Bloomberg, S&P500 stock prices	2006–2013	Price data, news, technical indicators	TGRU, stock2vec	Accuracy, precision, AUROC	Keras, Python
[169]	S&P500 Index, 15 stocks in S&P500	2006–2013	News from Reuters and Bloomberg	CNN	Accuracy, MCC	–
[170]	S&P500 index news from Reuters	2006–2013	Financial news titles, Technical indicators	SI-RCNN (LSTM + CNN)	Accuracy	–
[171]	10 stocks in Nikkei 225 and news	2001–2008	Textual information and Stock prices	Paragraph Vector + LSTM	Profit	–
[172]	NIFTY50 Index, NIFTY Bank/Auto/IT/Energy Index, News	2013–2017	Index data, news	LSTM	MCC, Accuracy	–
[173]	Price data, index data, news, social media data	2015	Price data, news from articles and social media	Coupled matrix and tensor	Accuracy, MCC	Jieba
[174]	HS300	2015–2017	Social media news, price data	RNN-Boost with LDA	Accuracy, MAE, MAPE, RMSE	Python, Scikit-learn

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Text mining studies without sentiment analysis for forecasting

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[175]	News and Chinese stock data	2014–2017	Selected words in a news	HAN	Accuracy, Annual return	–
[176]	News, stock prices from Hong Kong Stock Exchange	2001	Price data and TF-IDF from news	ELM, DLR, PCA, BELM, KELM, NN	Accuracy	Matlab
[177]	TWSE index, 4 stocks in TWSE	2001–2017	Technical indicators, Price data, News	CNN + LSTM	RMSE, Profit	Keras, Python, TALIB
[178]	Stock of Tsugami Corporation	2013	Price data	LSTM	RMSE	Keras, Tensorflow
[179]	News, Nikkei Stock Average and 10-Nikkei companies	1999–2008	news, MACD	RNN, RBM+DBN	Accuracy, *P*-value	–
[180]	ISMIS 2017 Data Mining Competition dataset	–	Expert identifier, classes	LSTM + GRU + FFNN	Accuracy	–
[181]	Reuters, Bloomberg News, S&P500 price	2006–2013	News and sentences	LSTM	Accuracy	–
[182]	APPL from S&P500 and news from Reuters	2011–2017	Input news, OCHLV, Technical indicators	CNN + LSTM, CNN+SVM	Accuracy, F1-score	Tensorflow
[183]	Nikkei225, S&P500, news from Reuters and Bloomberg	2001–2013	Stock price data and news	DGM	Accuracy, MCC, %profit	–
[184]	Stocks from S&P500	2006–2013	Text (news) and Price data	LAR+News, RF+News	MAPE, RMSE	–

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Financial sentiment studies coupled with text mining without forecasting

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[85]	883 BHC from EDGAR	2006–2017	Tokens, weighted sentiment polarity, leverage and ROA	CNN, LSTM, SVM, Random Forest	Accuracy, Precision, Recall, F1-score	Keras, Python, Scikit-learn
[185]	SemEval-2017 dataset, financial text, news, stock market data	2017	Sentiments in Tweets, News headlines	Ensemble SVR, CNN, LSTM, GRU	Cosine similarity score, agreement score, class score	Python, Keras, Scikit Learn
[186]	Financial news from Reuters	2006–2015	Word vector, Lexical and Contextual input	Targeted dependency tree LSTM	Cumulative abnormal return	–
[187]	Stock sentiment analysis from StockTwits	2015	StockTwits messages	LSTM, Doc2Vec, CNN	Accuracy, precision, recall, f-measure, AUC	–
[188]	Sina Weibo, Stock market records	2012–2015	Technical indicators, sentences	DRSE	F1-score, precision, recall, accuracy, AUROC	Python
[189]	News from NowNews, AppleDaily, LTN, MoneyDJ for 18 stocks	2013–2014	Text, Sentiment	LSTM, CNN	Return	Python, Tensorflow
[190]	StockTwits	2008–2016	Sentences, StockTwits messages	CNN, LSTM, GRU	MCC, WSURT	Keras, Tensorflow
[191]	Financial statements of Japan companies	–	Sentences, text	DMLP	Precision, recall, f-score	–
[192]	Twitter posts, news headlines	–	Sentences, text	Deep-FASP	Accuracy, MSE, R^2	–
[193]	Forums data	2004–2013	Sentences and keywords	Recursive neural tensor networks	Precision, recall, f-measure	–
[194]	News from Financial Times related US stocks	–	Sentiment of news headlines	SVR, Bidirectional LSTM	Cosine similarity	Python, Scikit Learn, Keras, Tensorflow

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications: Other text mining studies

Art.	Data set	Period	Feature set	Method	Performance criteria	**Env.**
[72]	News from NowNews, AppleDaily, LTN, MoneyDJ for 18 stocks	2013–2014	Text, Sentiment	DMLP	Return	Python, Tensorflow
[86]	The event data set for large European banks, news articles from Reuters	2007–2014	Word, sentence	DMLP +NLP preprocess	Relative usefulness, F1-score	–
[87]	Event dataset on European banks, news from Reuters	2007–2014	Text, sentence	Sentence vector + DFFN	Usefulness, F1-score, AUROC	–
[88]	News from Reuters, fundamental data	2007–2014	Financial ratios and news text	doc2vec + NN	Relative usefulness	Doc2vec
[121]	Real-world data for automobile insurance company labeled as fradulent	–	Car, insurance and accident related features	DMLP + LDA	TP, FP, Accuracy, Precision, F1-score	–
[123]	Financial transactions	–	Transaction data	LSTM	t-SNE	–
[195]	Taiwan's National Pension Insurance	2008–2014	Insured's id, area-code, gender, etc.	RNN	Accuracy, total error	Python
[196]	StockTwits	2015–2016	Sentences, StockTwits messages	Doc2vec, CNN	Accuracy, precision, recall, f-measure, AUC	Python, Tensorflow

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Algo-trading applications embedded with time series forecasting models

Art.	Data set	Period	Feature set	Method	Performance criteria	Environment
[33]	GarantiBank in BIST, Turkey	2016	OCHLV, Spread, Volatility, Turnover, etc.	PLR, Graves LSTM	MSE, RMSE, MAE, RSE, Correlation R-square	Spark
[34]	CSI300, Nifty50, HSI, Nikkei 225, S&P500, DJIA	2010–2016	OCHLV, Technical Indicators	WT, Stacked autoencoders, LSTM	MAPE, Correlation coefficient, THEIL-U	–
[35]	Chinese Stocks	2007–2017	OCHLV	CNN + LSTM	Annualized Return, Mxm Retracement	Python
[36]	50 stocks from NYSE	2007–2016	Price data	SFM	MSE	–
[37]	The LOB of 5 stocks of Finnish Stock Market	2010	FI-2010 dataset: bid/ask and volume	WMTR, MDA	Accuracy, Precision, Recall, F1-Score	–
[38]	300 stocks from SZSE, Commodity	2014–2015	Price data	FDDR, DMLP+RL	Profit, return, SR, profit-loss curves	Keras
[39]	S&P500 Index	1989–2005	Price data, Volume	LSTM	Return, STD, SR, Accuracy	Python, TensorFlow, Keras, R, H2O
[40]	Stock of National Bank of Greece (ETE).	2009–2014	FTSE100, DJIA, GDAX, NIKKEI225, EUR/USD, Gold	GASVR, LSTM	Return, volatility, SR, Accuracy	Tensorflow
[41]	Chinese stock-IF-IH-IC contract	2016–2017	Decisions for price change	MODRL+LSTM	Profit and loss, SR	–
[42]	Singapore Stock Market Index	2010–2017	OCHL of last 10 days of Index	DMLP	RMSE, MAPE, Profit, SR	–
[43]	GBP/USD	2017	Price data	Reinforcement Learning + LSTM + NES	SR, downside deviation ratio, total profit	Python, Keras, Tensorflow
[44]	Commodity, FX future, ETF	1991–2014	Price Data	DMLP	SR, capability ratio, return	C++, Python
[45]	USD/GBP, S&P500, FTSE100, oil, gold	2016	Price data	AE + CNN	SR, % volatility, avg return/trans, rate of return	H2O

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Algo-trading applications embedded with time series forecasting models

Art.	Data set	Period	Feature set	Method	Performance criteria	Environment
[46]	Bitcoin, Dash, Ripple, Monero, Litecoin, Dogecoin, Nxt, Namecoin	2014–2017	MA, BOLL, the CRIX returns, Euribor interest rates, OCHLV	LSTM, RNN, DMLP	Accuracy, F1-measure	Python, Tensorflow
[47]	S&P500, KOSPI, HSI, and EuroStoxx50	1987–2017	200-days stock price	Deep Q-Learning, DMLP	Total profit, Correlation	–
[48]	Stocks in the S&P500	1990–2015	Price data	DMLP, GBT, RF	Mean return, MDD, Calmar ratio	H2O
[49]	Fundamental and Technical Data, Economic Data	–	Fundamental , technical and market information	CNN	–	–

Deep learning for financial applications:

Classification (buy–sell signal, or trend detection) based algo-trading models

Art.	Data set	Period	Feature set	Method	Performance criteria	Environment
[51]	Stocks in Dow30	1997–2017	RSI	DMLP with genetic algorithm	Annualized return	Spark MLlib, Java
[52]	SPY ETF, 10 stocks from S&P500	2014–2016	Price data	FFNN	Cumulative gain	MatConvNet, Matlab
[53]	Dow30 stocks	2012–2016	Close data and several technical indicators	LSTM	Accuracy	Python, Keras, Tensorflow, TALIB
[54]	High-frequency record of all orders	2014–2017	Price data, record of all orders, transactions	LSTM	Accuracy	–
[55]	Nasdaq Nordic (Kesko Oyj, Outokumpu Oyj, Sampo, Rautaruukki, Wartsila Oyj)	2010	Price and volume data in LOB	LSTM	Precision, Recall, F1-score, Cohen's k	–
[56]	17 ETFs	2000–2016	Price data, technical indicators	CNN	Accuracy, MSE, Profit, AUROC	Keras, Tensorflow
[57]	Stocks in Dow30 and 9 Top Volume ETFs	1997–2017	Price data, technical indicators	CNN with feature imaging	Recall, precision, F1-score, annualized return	Python, Keras, Tensorflow, Java
[58]	FTSE100	2000–2017	Price data	CAE	TR, SR, MDD, mean return	–
[59]	Nasdaq Nordic (Kesko Oyj, Outokumpu Oyj, Sampo, Rautaruukki, Wartsila Oyj)	2010	Price, Volume data, 10 orders of the LOB	CNN	Precision, Recall, F1-score, Cohen's k	Theano, Scikit learn, Python
[60]	Borsa Istanbul 100 Stocks	2011–2015	75 technical indicators and OCHLV	CNN	Accuracy	Keras
[61]	ETFs and Dow30	1997–2007	Price data	CNN with feature imaging	Annualized return	Keras, Tensorflow
[62]	8 experimental assets from bond/derivative market	–	Asset prices data	RL, DMLP, Genetic Algorithm	Learning and genetic algorithm error	–
[63]	10 stocks from S&P500	–	Stock Prices	TDNN, RNN, PNN	Missed opportunities, false alarms ratio	–
[64]	London Stock Exchange	2007–2008	Limit order book state, trades, buy/sell orders, order deletions	CNN	Accuracy, kappa	Caffe
[65]	Cryptocurrencies, Bitcoin	2014–2017	Price data	CNN, RNN, LSTM	Accumulative portfolio value, MDD, SR	–

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:
Stand-alone and/or other algorithmic models

Art.	Data set	Period	Feature set	Method	Performance criteria	Environment
[66]	DAX, FTSE100, call/put options	1991–1998	Price data	Markov model, RNN	Ewa-measure, iv, daily profits' mean and std	–
[67]	Taiwan Stock Index Futures, Mini Index Futures	2012–2014	Price data to image	Visualization method + CNN	Accumulated profits,accuracy	–
[68]	Energy-Sector/ Company-Centric Tweets in S&P500	2015–2016	Text and Price data	LSTM, RNN, GRU	Return, SR, precision, recall, accuracy	Python, Tweepy API
[69]	CME FIX message	2016	Limit order book, time-stamp, price data	RNN	Precision, recall, F1-measure	Python, TensorFlow, R
[70]	Taiwan stock index futures (TAIFEX)	2017	Price data	Agent based RL with CNN pre-trained	Accuracy	–
[71]	Stocks from S&P500	2010–2016	OCHLV	DCNL	PCC, DTW, VWL	Pytorch
[72]	News from NowNews, AppleDaily, LTN, MoneyDJ for 18 stocks	2013–2014	Text, Sentiment	DMLP	Return	Python, Tensorflow
[73]	489 stocks from S&P500 and NASDAQ-100	2014–2015	Limit Order Book	Spatial neural network	Cross entropy error	NVIDIA's cuDNN
[74]	Experimental dataset	–	Price data	DRL with CNN, LSTM, GRU, DMLP	Mean profit	Python

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Credit scoring or classification studies

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[77]	The XR 14 CDS contracts	2016	Recovery rate, spreads, sector and region	DBN+RBM	AUROC, FN, FP, Accuracy	WEKA
[78]	German, Japanese credit datasets	–	Personal financial variables	SVM + DBN	Weighted-accuracy, TP, TN	–
[79]	Credit data from Kaggle	–	Personal financial variables	DMLP	Accuracy, TP, TN, G-mean	–
[80]	Australian, German credit data	–	Personal financial variables	GP + AE as Boosted DMLP	FP	Python, Scikit-learn
[81]	German, Australian credit dataset	–	Personal financial variables	DCNN, DMLP	Accuracy, False/Missed alarm	–
[82]	Consumer credit data from Chinese finance company	–	Relief algorithm chose the 50 most important features	CNN + Relief	AUROC, K-s statistic, Accuracy	Keras
[83]	Credit approval dataset by UCI Machine Learning repo	–	UCI credit approval dataset	Rectifier, Tanh, Maxout DL	–	AWS EC2, H2O, R

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Financial distress, bankruptcy, bank risk, mortgage risk, crisis forecasting studies.

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[84]	966 french firms	–	Financial ratios	RBM+SVM	Precision, Recall	–
[85]	883 BHC from EDGAR	2006–2017	Tokens, weighted sentiment polarity, leverage and ROA	CNN, LSTM, SVM, RF	Accuracy, Precision, Recall, F1-score	Keras, Python, Scikit-learn
[86]	The event data set for large European banks, news articles from Reuters	2007–2014	Word, sentence	DMLP +NLP preprocess	Relative usefulness, F1-score	–
[87]	Event dataset on European banks, news from Reuters	2007–2014	Text, sentence	Sentence vector + DFFN	Usefulness, F1-score, AUROC	–
[88]	News from Reuters, fundamental data	2007–2014	Financial ratios and news text	doc2vec + NN	Relative usefulness	Doc2vec
[89]	Macro/Micro economic variables, Bank characteristics/performance variables from BHC	1976–2017	Macro economic variables and bank performances	CGAN, MVN, MV-t, LSTM, VAR, FE-QAR	RMSE, Log likelihood, Loan loss rate	–
[90]	Financial statements of French companies	2002–2006	Financial ratios	DBN	Recall, Precision, F1-score, FP, FN	–
[91]	Stock returns of American publicly-traded companies from CRSP	2001–2011	Price data	DBN	Accuracy	Python, Theano
[92]	Financial statements of several companies from Japanese stock market	2002–2016	Financial ratios	CNN	F1-score, AUROC	–
[93]	Mortgage dataset with local and national economic factors	1995–2014	Mortgage related features	DMLP	Negative average log-likelihood	AWS
[94]	Mortgage data from Norwegian financial service group, DNB	2012–2016	Personal financial variables	CNN	Accuracy, Sensitivity, Specificity, AUROC	–
[95]	Private brokerage company's real data of risky transactions	–	250 features: order details, etc.	CNN, LSTM	F1-Score	Keras, Tensorflow
[96]	Several datasets combined to create a new one	1996–2017	Index data, 10-year Bond yield, exchange rates,	Logit, CART, RF, SVM, NN, XGBoost, DMLP	AUROC, KS, G-mean, likelihood ratio, DP, BA, WBA	R

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:
Fraud detection studies

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[114]	Debit card transactions by a local Indonesia bank	2016–2017	Financial transaction amount on several time periods	CNN, Stacked-LSTM, CNN-LSTM	AUROC	–
[115]	Credit card transactions from retail banking	2017	Transaction variables and several derived features	LSTM, GRU	Accuracy	Keras
[116]	Card purchases' transactions	2014–2015	Probability of fraud per currency/origin country, other fraud related features	DMLP	AUROC	–
[117]	Transactions made with credit cards by European cardholders	2013	Personal financial variables to PCA	DMLP, RF	Recall, Precision, Accuracy	–
[118]	Credit-card transactions	2015	Transaction and bank features	LSTM	AUROC	Keras, Scikit-learn
[119]	Databases of foreign trade of the Secretariat of Federal Revenue of Brazil	2014	8 Features: Foreign Trade, Tax, Transactions, Employees, Invoices, etc	AE	MSE	H2O, R
[120]	Chamber of Deputies open data, Companies data from Secretariat of Federal Revenue of Brazil	2009–2017	21 features: Brazilian State expense, party name, Type of expense, etc.	Deep Autoencoders	MSE, RMSE	H2O, R
[121]	Real-world data for automobile insurance company labeled as fradulent	–	Car, insurance and accident related features	DMLP + LDA	TP, FP, Accuracy, Precision, F1-score	–
[122]	Transactions from a giant online payment platform	2006	Personal financial variables	GBDT+DMLP	AUROC	–
[123]	Financial transactions	–	Transaction data	LSTM	t-SNE	–
[124]	Empirical data from Greek firms	–	–	DQL	Revenue	Torch

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications: Portfolio management studies

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[65]	Cryptocurrencies, Bitcoin	2014–2017	Price data	CNN, RNN, LSTM	Accumulative portfolio value, MDD, SR	–
[127]	Stocks from NYSE, AMEX, NASDAQ	1965–2009	Price data	Autoencoder + RBM	Accuracy, confusion matrix	–
[128]	20 stocks from S&P500	2012–2015	Technical indicators	DMLP	Accuracy	Python, Scikit Learn, Keras, Theano
[129]	Chinese stock data	2012–2013	Technical, fundamental data	Logistic Regression, RF, DMLP	AUC, accuracy, precision, recall, f1, tpr, fpr	Keras, Tensorflow, Python, Scikit learn
[130]	Top 5 companies in S&P500	–	Price data and Financial ratios	LSTM, Auto-encoding, Smart indexing	CAGR	–
[131]	IBB biotechnology index, stocks	2012–2016	Price data	Auto-encoding, Calibrating, Validating, Verifying	Returns	–
[132]	Taiwans stock market	–	Price data	Elman RNN	MSE, return	–
[133]	FOREX (EUR/USD, etc.), Gold	2013	Price data	Evolino RNN	Return	Python
[134]	Stocks in NYSE, AMEX, NASDAQ, TAQ intraday trade	1993–2017	Price, 15 firm characteristics	LSTM+DMLP	Monthly return, SR	Python,Keras, Tensorflow in AWS
[135]	S&P500	1985–2006	monthly and daily log-returns	DBN+MLP	Validation, Test Error	Theano, Python, Matlab
[136]	10 stocks in S&P500	1997–2016	OCHLV, Price data	RNN, LSTM, GRU	Accuracy, Monthly return	Keras, Tensorflow
[137]	Analyst reports on the TSE and Osaka Exchange	2016–2018	Text	LSTM, CNN, Bi-LSTM	Accuracy, R^2	R, Python, MeCab
[138]	Stocks from Chinese/American stock market	2015–2018	OCHLV, Fundamental data	DDPG, PPO	SR, MDD	–
[139]	Hedge fund monthly return data	1996–2015	Return, SR, STD, Skewness, Kurtosis, Omega ratio, Fund alpha	DMLP	Sharpe ratio, Annual return, Cum. return	–
[140]	12 most-volumed cryptocurrency	2015–2016	Price data	CNN + RL	SR, portfolio value, MDD	–

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:
Asset pricing and derivatives market studies

Art.	Der. type	Data set	Period	Feature set	Method	Performance criteria	Env.
[137]	Asset pricing	Analyst reports on the TSE and Osaka Exchange	2016–2018	Text	LSTM, CNN, Bi-LSTM	Accuracy, R^2	R, Python, MeCab
[142]	Options	Simulated a range of call option prices	–	Price data, option strike/maturity, dividend/risk free rates, volatility	DMLP	RMSE, the average percentage pricing error	Tensorflow
[143]	Futures, Options	TAIEX Options	2017	OCHLV, fundamental analysis, option price	DMLP, DMLP with Black scholes	RMSE, MAE, MAPE	–
[144]	Equity returns	Returns in NYSE, AMEX, NASDAQ	1975–2017	57 firm characteristics	Fama–French n-factor model DL	R^2,RMSE	Tensorflow

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Cryptocurrency and blockchain studies

Art.	Data set	Period	Feature set	Method	Performance criteria	Env.
[46]	Bitcoin, Dash, Ripple, Monero, Litecoin, Dogecoin, Nxt, Namecoin	2014–2017	MA, BOLL, the CRIX daily returns, Euribor interest rates, OCHLV of EURO/UK, EURO/USD, US/JPY	LSTM, RNN, DMLP	Accuracy, F1-measure	Python, Tensorflow
[65]	Cryptocurrencies, Bitcoin	2014–2017	Price data	CNN	Accumulative portfolio value, MDD, SR	–
[140]	12 most-volumed cryptocurrency	2015–2016	Price data	CNN + RL	SR, portfolio value, MDD	
[145]	Bitcoin data	2010–2017	Hash value, bitcoin address, public/private key, digital signature, etc.	Takagi–Sugeno Fuzzy cognitive maps	Analytical hierarchy process	–
[146]	Bitcoin data	2012, 2013, 2016	TransactionId, input/output Addresses, timestamp	Graph embedding using heuristic, laplacian eigen-map, deep AE	F1-score	–
[147]	Bitcoin, Litecoin, StockTwits	2015–2018	OCHLV, technical indicators, sentiment analysis	CNN, LSTM, State Frequency Model	MSE	Keras, Tensorflow
[148]	Bitcoin	2013–2016	Price data	Bayesian optimized RNN, LSTM	Sensitivity, specificity, precision, accuracy, RMSE	Keras, Python, Hyperas

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:

Other theoretical or conceptual studies

Art.	SubTopic	IsTimeSeries?	Data set	Period	Feature set	Method
[197]	Analysis of AE, SVD	Yes	Selected stocks from the IBB index and stock of Amgen Inc.	2012–2014	Price data	AE, SVD
[198]	Fraud Detection in Banking	No	Risk Management / Fraud Detection	–	–	DRL

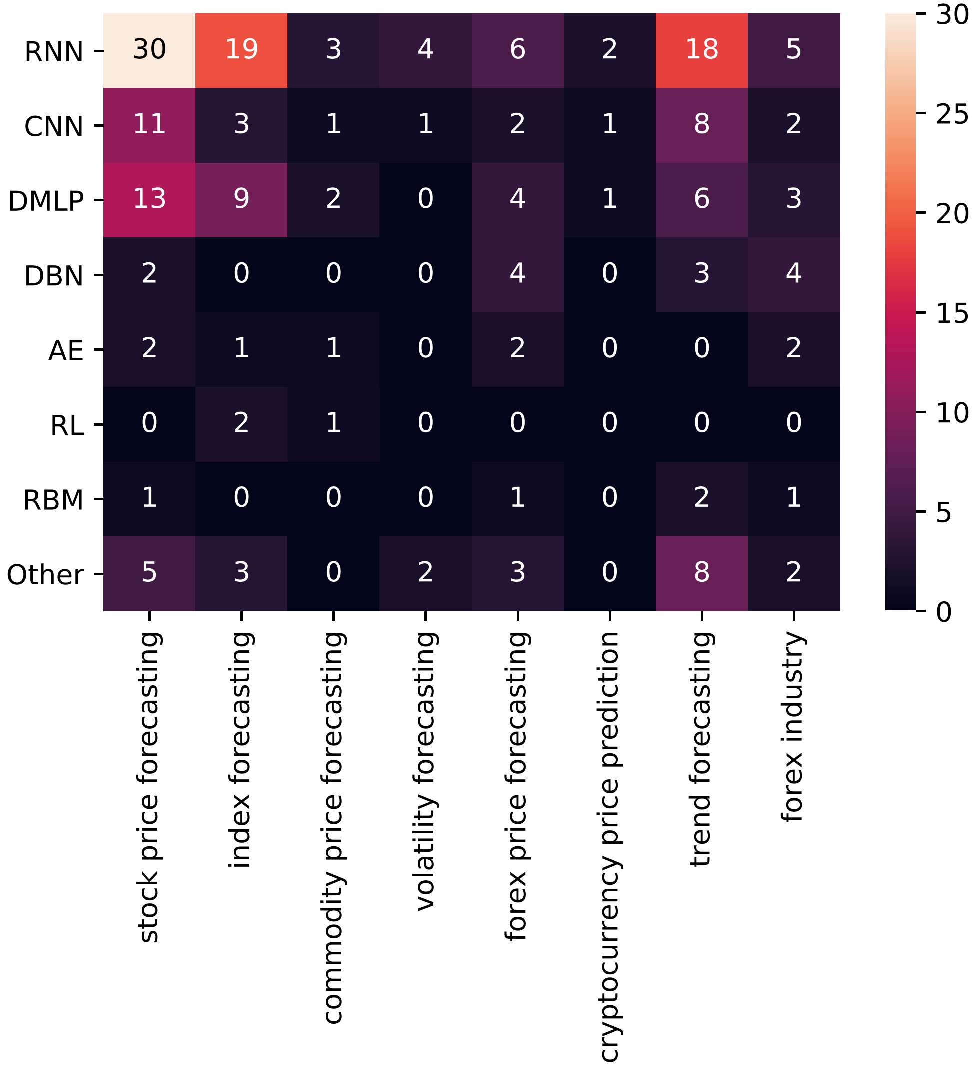
Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Deep learning for financial applications:
Other financial applications

Art.	Subtopic	Data set	Period	Feature set	Method	Performance criteria	Env.
[47]	Improving trading decisions	S&P500, KOSPI, HSI, and EuroStoxx50	1987–2017	200-days stock price	Deep Q-Learning and DMLP	Total profit, Correlation	–
[193]	Identifying Top Sellers In Underground Economy	Forums data	2004–2013	Sentences and keywords	Recursive neural tensor networks	Precision, recall, f-measure	–
[195]	Predicting Social Ins. Payment Behavior	Taiwan's National Pension Insurance	2008–2014	Insured's id, area-code, gender, etc.	RNN	Accuracy, total error	Python
[199]	Speedup	45 CME listed commodity and FX futures	1991–2014	Price data	DNN	–	–
[200]	Forecasting Fundamentals	Stocks in NYSE, NASDAQ or AMEX exchanges	1970–2017	16 fundamental features from balance sheet	DMLP, LFM	MSE, Compound annual return, SR	–
[201]	Predicting Bank Telemarketing	Phone calls of bank marketing data	2008–2010	16 finance-related attributes	CNN	Accuracy	–
[202]	Corporate Performance Prediction	22 pharmaceutical companies data in US stock market	2000–2015	11 financial and 4 patent indicator	RBM, DBN	RMSE, profit	–

Source: Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.

Financial time series forecasting with deep learning: Topic-model heatmap



Source: Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu (2020), "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019." Applied Soft Computing 90 (2020): 106181.

Stock price forecasting using only raw time series data

Art.	Data set	Period	Feature set	Lag	Horizon	Method	Performance criteria	Env.
[80]	38 stocks in KOSPI	2010–2014	Lagged stock returns	50 min	5 min	DNN	NMSE, RMSE, MAE, MI	–
[81]	China stock market, 3049 Stocks	1990–2015	OCHLV	30 d	3 d	LSTM	Accuracy	Theano, Keras
[82]	Daily returns of 'BRD' stock in Romanian Market	2001–2016	OCHLV	–	1 d	LSTM	RMSE, MAE	Python, Theano
[83]	297 listed companies of CSE	2012–2013	OCHLV	2 d	1 d	LSTM, SRNN, GRU	MAD, MAPE	Keras
[84]	5 stock in NSE	1997–2016	OCHLV, Price data, turnover and number of trades.	200 d	1..10 d	LSTM, RNN, CNN, MLP	MAPE	–
[85]	Stocks of Infosys, TCS and CIPLA from NSE	2014	Price data	–	–	RNN, LSTM and CNN	Accuracy	–
[86]	10 stocks in S&P500	1997–2016	OCHLV, Price data	36 m	1 m	RNN, LSTM, GRU	Accuracy, Monthly return	Keras, Tensorflow
[87]	Stocks data from S&P500	2011–2016	OCHLV	1 d	1 d	DBN	MSE, norm-RMSE, MAE	–
[88]	High-frequency transaction data of the CSI300 futures	2017	Price data	–	1 min	DNN, ELM, RBF	RMSE, MAPE, Accuracy	Matlab
[89]	Stocks in the S&P500	1990–2015	Price data	240 d	1 d	DNN, GBT, RF	Mean return, MDD, Calmar ratio	H2O
[90]	ACI Worldwide, Staples, and Seagate in NASDAQ	2006–2010	Daily closing prices	17 d	1 d	RNN, ANN	RMSE	–
[91]	Chinese Stocks	2007–2017	OCHLV	30 d	1..5 d	CNN + LSTM	Annualized Return, Mxm Retracement	Python
[92]	20 stocks in S&P500	2010–2015	Price data	–	–	AE + LSTM	Weekly Returns	–
[93]	S&P500	1985–2006	Monthly and daily log-returns	*	1 d	DBN+MLP	Validation, Test Error	Theano, Python, Matlab
[94]	12 stocks from SSE Composite Index	2000–2017	OCHLV	60 d	1..7 d	DWNN	MSE	Tensorflow
[95]	50 stocks from NYSE	2007–2016	Price data	–	1d, 3 d, 5 d	SFM	MSE	–

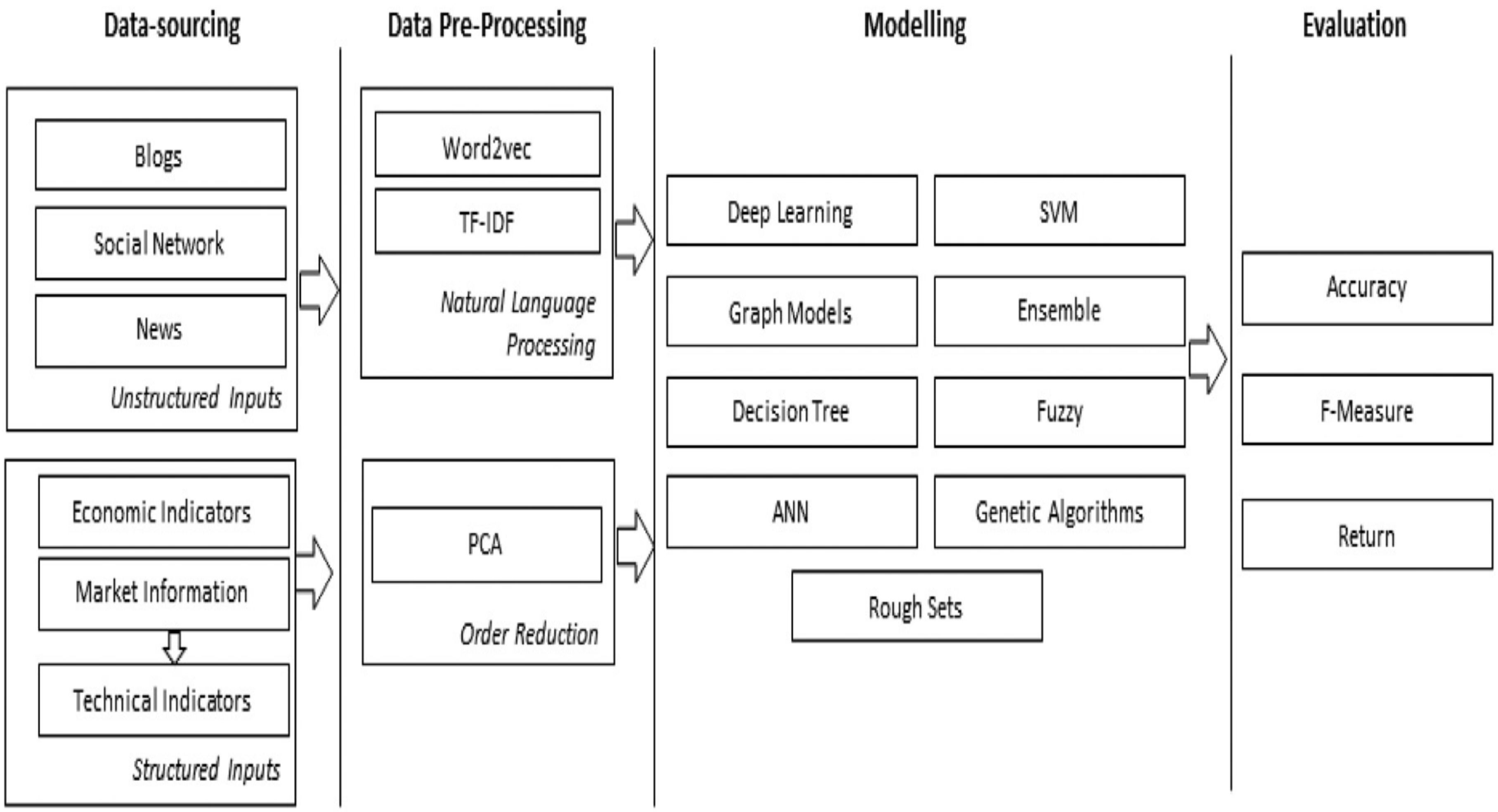
Source: Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu (2020), "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019." Applied Soft Computing 90 (2020): 106181.

Stock price forecasting using various data

Art.	Data set	Period	Feature set	Lag	Horizon	Method	Performance criteria	Env.
[96]	Japan Index constituents from WorldScope	1990–2016	25 Fundamental Features	10 d	1 d	DNN	Correlation, Accuracy, MSE	Tensorflow
[97]	Return of S&P500	1926–2016	Fundamental Features:	–	1 s	DNN	MSPE	Tensorflow
[98]	U.S. low-level disaggregated macroeconomic time series	1959–2008	GDP, Unemployment rate, Inventories, etc.	–	–	DNN	R^2	–
[99]	CDAX stock market data	2010–2013	Financial news, stock market data	20 d	1 d	LSTM	MSE, RMSE, MAE, Accuracy, AUC	TensorFlow, Theano, Python, Scikit-Learn
[100]	Stock of Tsugami Corporation	2013	Price data	–	–	LSTM	RMSE	Keras, Tensorflow
[101]	Stocks in China's A-share	2006–2007	11 technical indicators	–	1 d	LSTM	AR, IR, IC	–
[102]	SCI prices	2008–2015	OCHL of change rate, price	7 d	–	EmotionalAnalysis + LSTM	MSE	–
[103]	10 stocks in Nikkei 225 and news	2001–2008	Textual information and Stock prices	10 d	–	Paragraph Vector + LSTM	Profit	–
[104]	TKC stock in NYSE and QQQQ ETF	1999–2006	Technical indicators, Price	50 d	1 d	RNN (Jordan–Elman)	Profit, MSE	Java
[105]	10 Stocks in NYSE	–	Price data, Technical indicators	20 min	1 min	LSTM, MLP	RMSE	–
[106]	42 stocks in China's SSE	2016	OCHLV, Technical Indicators	242 min	1 min	GAN (LSTM, CNN)	RMSRE, DPA, GAN-F, GAN-D	–
[107]	Google's daily stock data	2004–2015	OCHLV, Technical indicators	20 d	1 d	$(2D)^2$ PCA + DNN	SMAPE, PCD, MAPE, RMSE, HR, TR, R^2	R, Matlab
[108]	GarantiBank in BIST, Turkey	2016	OCHLV, Volatility, etc.	–	–	PLR, Graves LSTM	MSE, RMSE, MAE, RSE, R^2	Spark
[109]	Stocks in NYSE, AMEX, NASDAQ, TAQ intraday trade	1993–2017	Price, 15 firm characteristics	80 d	1 d	LSTM+MLP	Monthly return, SR	Python,Keras, Tensorflow in AWS
[110]	Private brokerage company's real data of risky transactions	–	250 features: order details, etc.	–	–	CNN, LSTM	F1-Score	Keras, Tensorflow
[111]	Fundamental and Technical Data, Economic Data	–	Fundamental , technical and market information	–	–	CNN	–	–
[112]	The LOB of 5 stocks of Finnish Stock Market	2010	FI-2010 dataset: bid/ask and volume	–	*	WMTR, MDA	Accuracy, Precision, Recall, F1-Score	–
[113]	Returns in NYSE, AMEX, NASDAQ	1975–2017	57 firm characteristics	*	–	Fama–French n-factor model DL	R^2, RMSE	Tensorflow

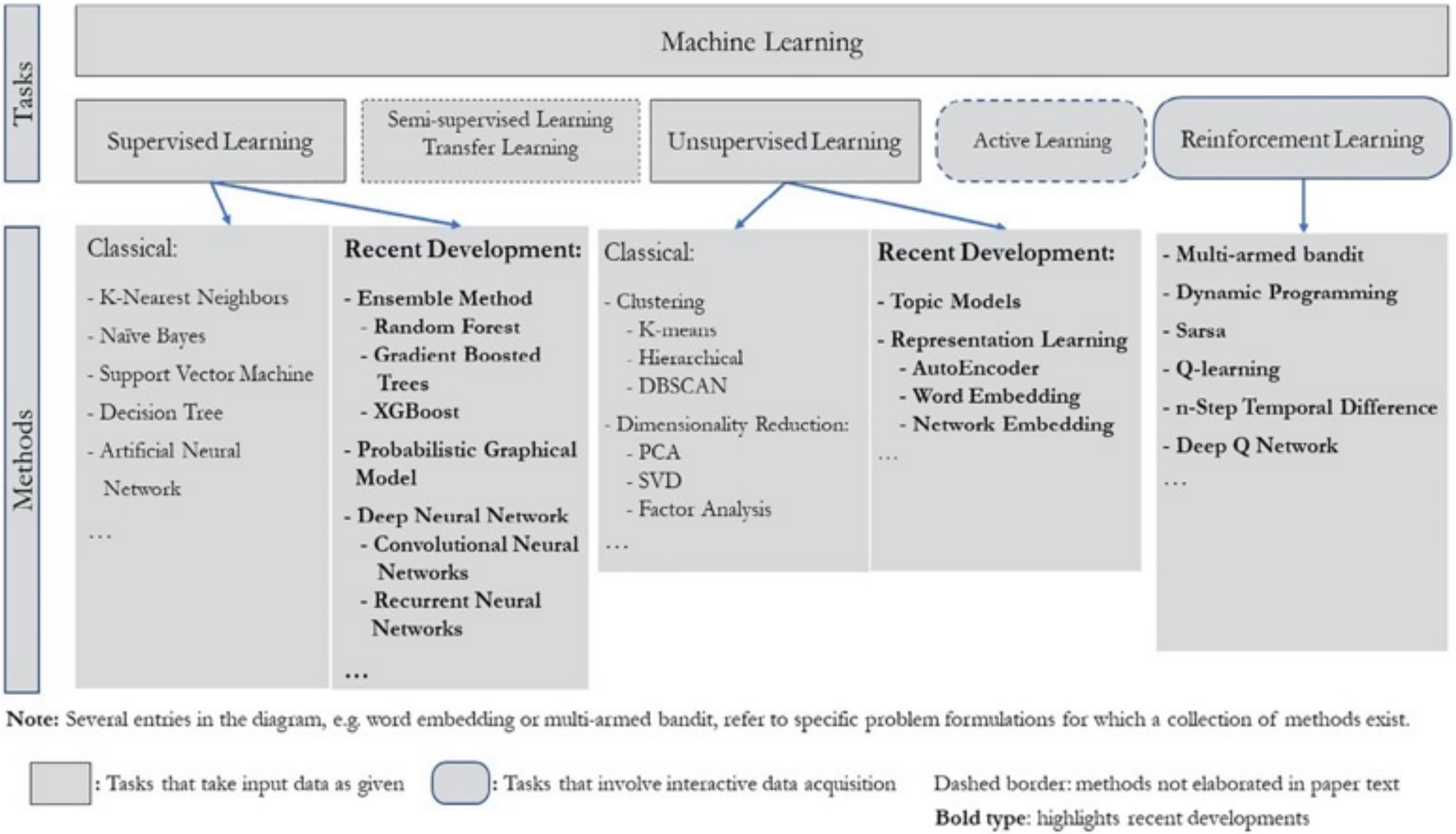
Source: Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu (2020), "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019." Applied Soft Computing 90 (2020): 106181.

Stock Market Movement Forecast:
Phases of the stock market modeling



Source: O. Bustos and A. Pomares-Quimbaya (2020), "Stock Market Movement Forecast: A Systematic Review." Expert Systems with Applications (2020): 113464.

Machine Learning Tasks and Methods



Source: Liye Ma and Baohong Sun (2020), "Machine learning and AI in marketing – Connecting computing power to human insights." International Journal of Research in Marketing, 37, no. 3, 481-504.

Decentralized Finance (DeFi)

Block Chain FinTech

Decentralized Finance (DeFi)

- A **global, open alternative** to the current **financial system**.
- Products that let you **borrow, save, invest, trade**, and more.
- Based on **open-source technology** that anyone can program with.

Source: https://ethereum.org/en/defi/

Traditional Finance
Centralized Finance (CeFi)

- Some people aren't granted access to set up a bank account or use financial services.
- Lack of access to financial services can prevent people from being employable.
- Financial services can block you from getting paid.
- A hidden charge of financial services is your personal data.
- Governments and centralized institutions can close down markets at will.
- Trading hours often limited to business hours of specific time zone.
- Money transfers can take days due to internal human processes.
- There's a premium to financial services because intermediary institutions need their cut.

Source: https://ethereum.org/en/defi/

DeFi vs. CeFi

Decentralized Finance (DeFi)	Traditional Finance (Centralized Finance; CeFi)
You hold your money.	Your money is held by companies.
You control where your money goes and how it's spent.	You have to trust companies not to mismanage your money, like lend to risky borrowers.
Transfers of funds happen in minutes.	Payments can take days due to manual processes.
Transaction activity is pseudonymous.	Financial activity is tightly coupled with your identity.
DeFi is open to anyone.	You must apply to use financial services.
The markets are always open.	Markets close because employees need breaks.
It's built on transparency – anyone can look at a product's data and inspect how the system works.	Financial institutions are closed books: you can't ask to see their loan history, a record of their managed assets, and so on.

Source: https://ethereum.org/en/defi/

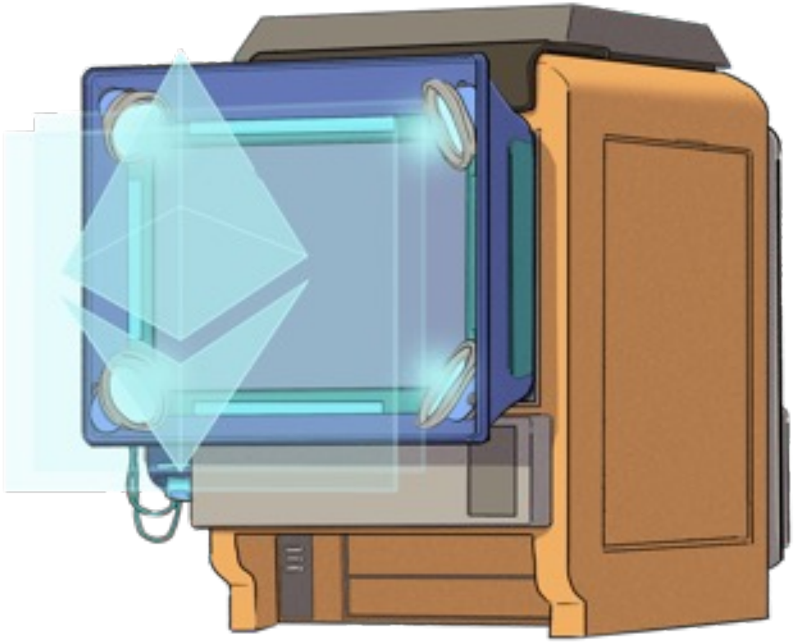
(DeFi)
Decentralized Applications (Dapps)

- **Ethereum-powered tools and services**
- **Dapps are a growing movement of applications that use Ethereum to disrupt business models or invent new ones**

Source: https://ethereum.org/en/defi/

The Internet of Assets

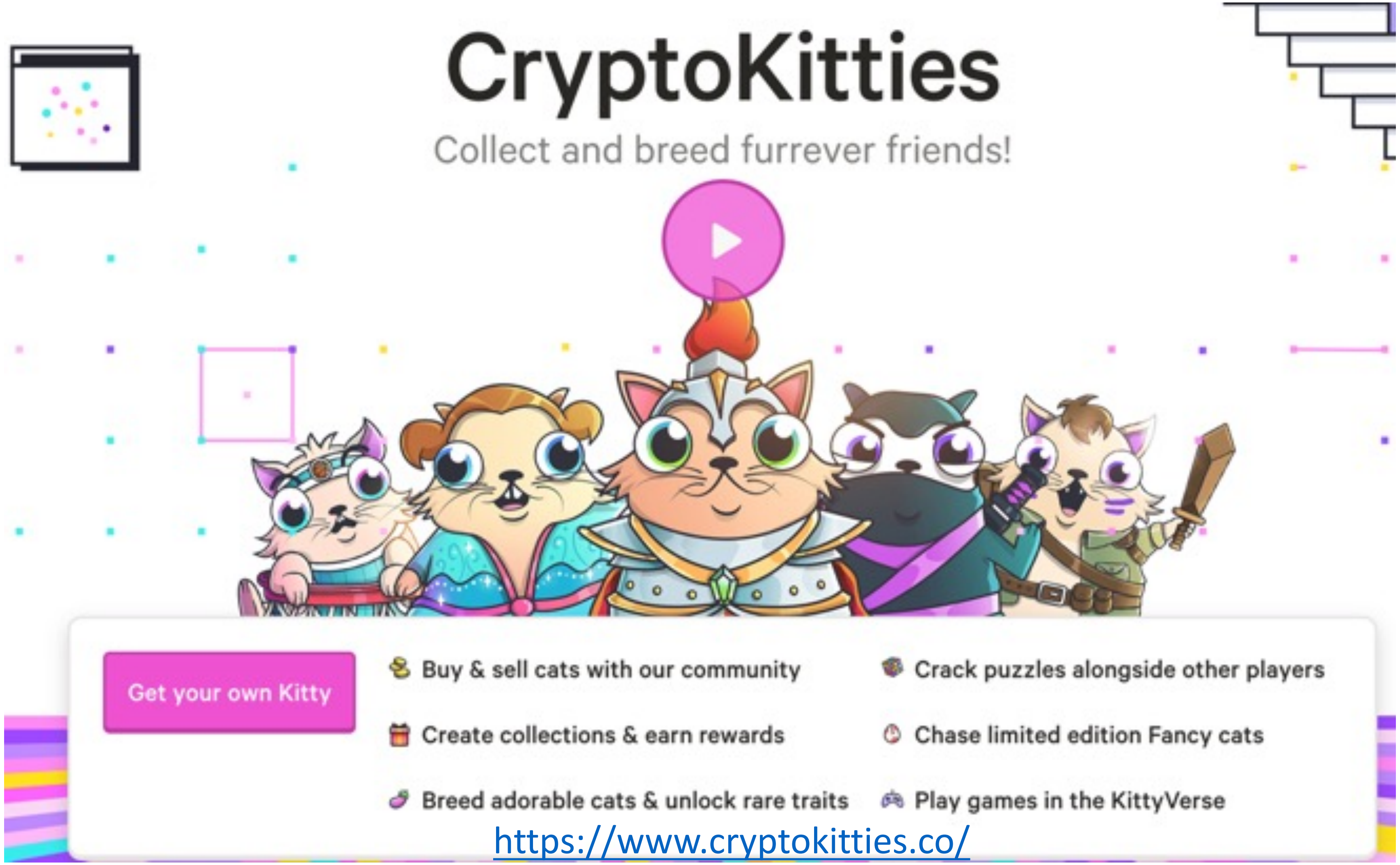
- **Ethereum** isn't just for **digital money**.
- Anything you can own can be **represented**, **traded** and **put to use** as **non-fungible tokens (NFTs)**.



Source: https://ethereum.org/en/defi/

Non-Fungible Tokens (NFT)

CryptoKitties



https://www.cryptokitties.co/

Source: Matt Fortnow and QuHarrison Terry (2021), The NFT Handbook - How to Create, Sell and Buy Non-Fungible Tokens, Wiley

Top Stablecoins

(Tether USDT, USD Coin USDC, Dai)

Digital money for everyday use

Stablecoins are Ethereum tokens designed to stay at a fixed value, even when the price of ETH changes.

CURRENCY	MARKET CAPITALIZATION	COLLATERAL TYPE
Tether	$69,136,810,713	Fiat
USD Coin	$32,359,142,012	Fiat
Binance USD	$13,083,174,132	Fiat
Dai	$6,265,852,093	Crypto
TrueUSD	$1,347,100,594	Fiat
PAX Gold	$318,953,291	Precious metals
HUSD	$296,254,105	Fiat
Gemini Dollar	$231,786,547	Fiat

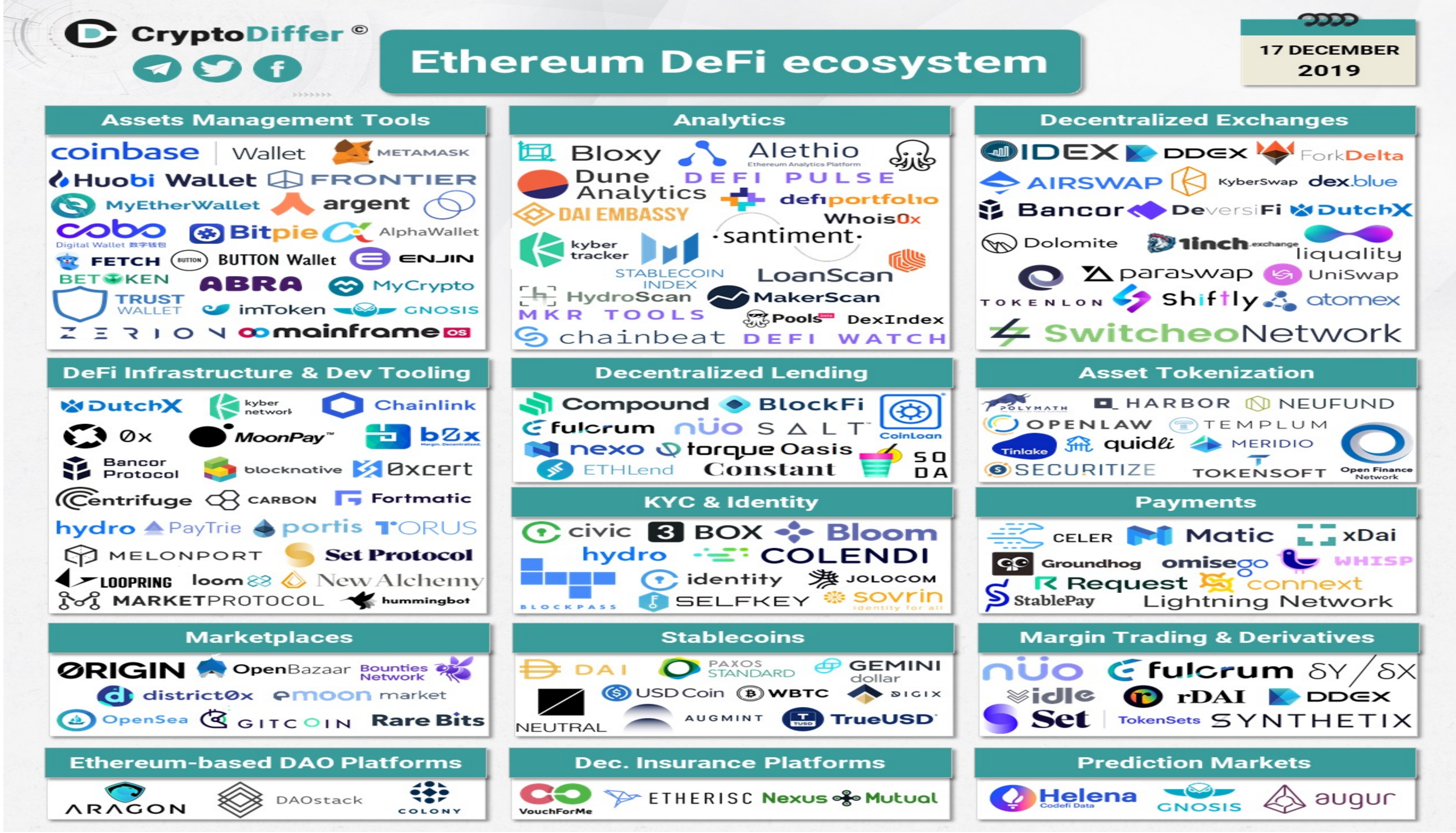
Source: https://ethereum.org/en/stablecoins/

Financial Stability Challenges

Crypto Ecosystem	• Operational, cyber, and governance risks • Integrity (market and AML/CFT) (Anti–Money Laundering / Combating the Financing of Terrorism) • Data availability / reliability • Challenges from cross-boarder activites
Stablecoins	• How stable are stablecoins? • Domestic and global regulatory and supervisory approaches
Macro-Financial	• Cryptoization, capital flows, and restrictions • Monetary policy transmission • Bank disintermediation

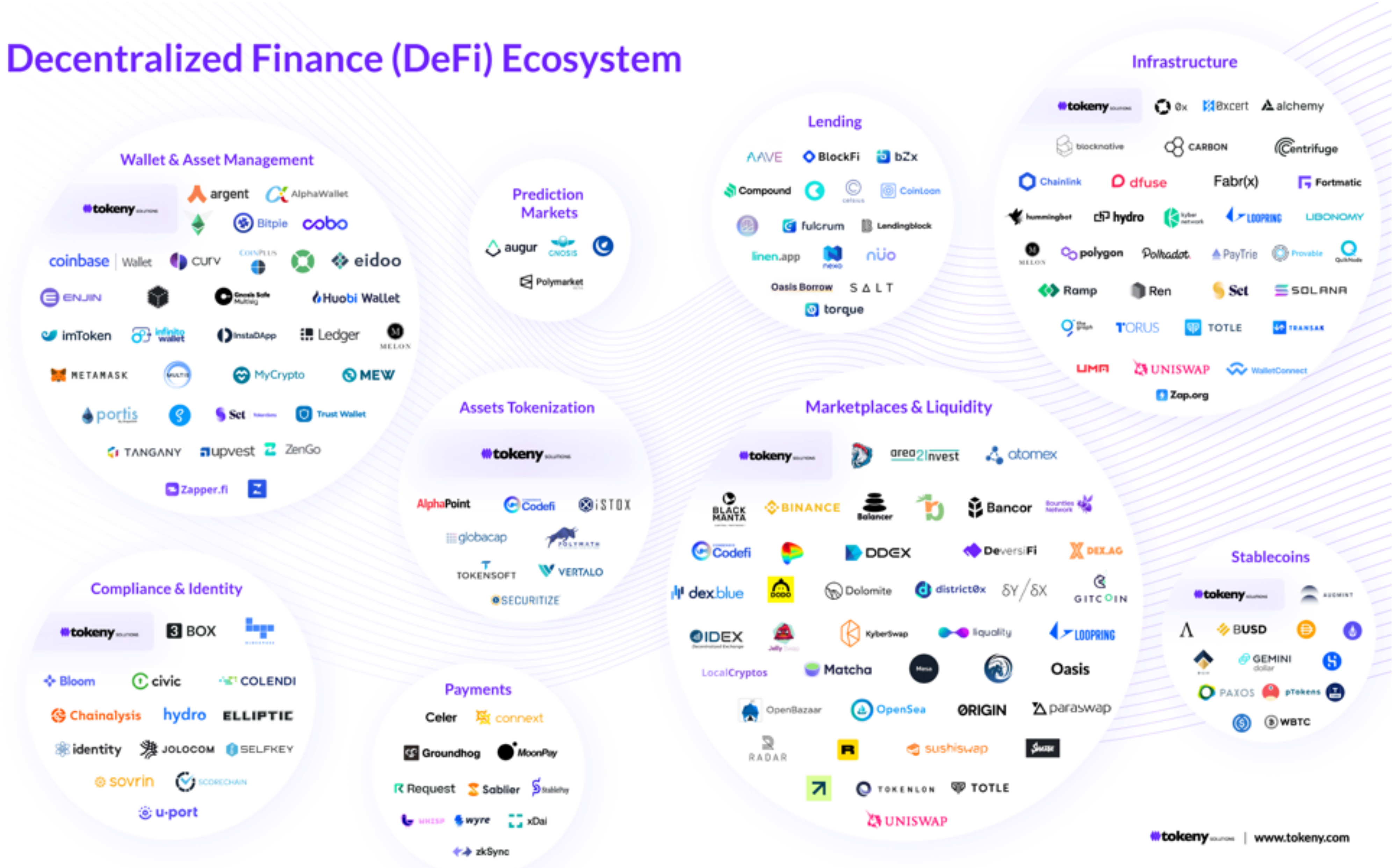
Source: Parma Bains, Mohamed Diaby, Dimitris Drakopoulos, Julia Faltermeier, Federico Grinberg, Evan Papageorgiou, Dmitri Petrov, Patrick Schneider, and Nobu Sugimoto (2021), The Crypto Ecosystem and Financial Stability Challenges, International Monetary Fund, October 2021

Ethereum DeFi Ecosystem



Source: https://cryptodiffer.com/news/ethereum-defi-ecosystem/

Decentralized Finance (DeFi) Ecosystem



Source: https://tokeny.com/defi-ecosystem/

Outline

- AI for Text Analytics
  - Natural Language Processing with Transformers: Building Language Applications with Hugging Face
  - Practical Natural Language Processing
- FinTech: Financial Services Innovation
- Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech

Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech

Time: November 8, 2021, 14:00 – 15:40 [Amsterdam Time GMT+2] MSNDS 2021 (Room:Tehran)

Min-Yuh Day, Ph.D,
Associate Professor

Graduate Institute of Information Management,
National Taipei University, Taiwan

https://web.ntpu.edu.tw/~myday

Knowledge Graph (KG)

- **Knowledge Graph (KG)**
  - **A knowledge graph is a multi-relational graph composed of entities and relations, which are regarded as nodes and different types of edges, respectively (Ji et al., 2021).**
  - **Represents knowledge as concepts (entities) and their relationships (Facts)**
  - **Triple of facts**
    - *SPO: (subject, predicate, object)*
    - *HRT: (head, relation, tail)*
- **Common Knowledge Graph: DBpedia, YAGO, Wikidata**

Knowledge Graph, Facts, Triple, Embedding

- G
    - Knowledge graph
- F
    - Set of facts
- (h, r, t)
    - Triple of head, relation, and tail
- $(\mathbf{h}, \mathbf{r}, \mathbf{t})$
    - Embedding of head, relation, and tail

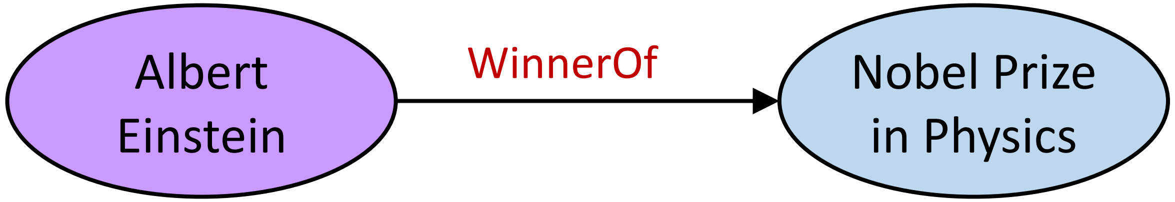
Source: Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems.

Knowledge Representation
Factual Triple and Knowledge Graph

- Albert Einstein, winner of the 1921 Nobel prize in physics
- The Nobel Prize in Physics 1921 was awarded to Albert Einstein "for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect."

Triple (Albert Einstein, **WinnerOf**, Nobel Prize in Physics)

Knowledge Graph



Factual **Triples** in Knowledge Base

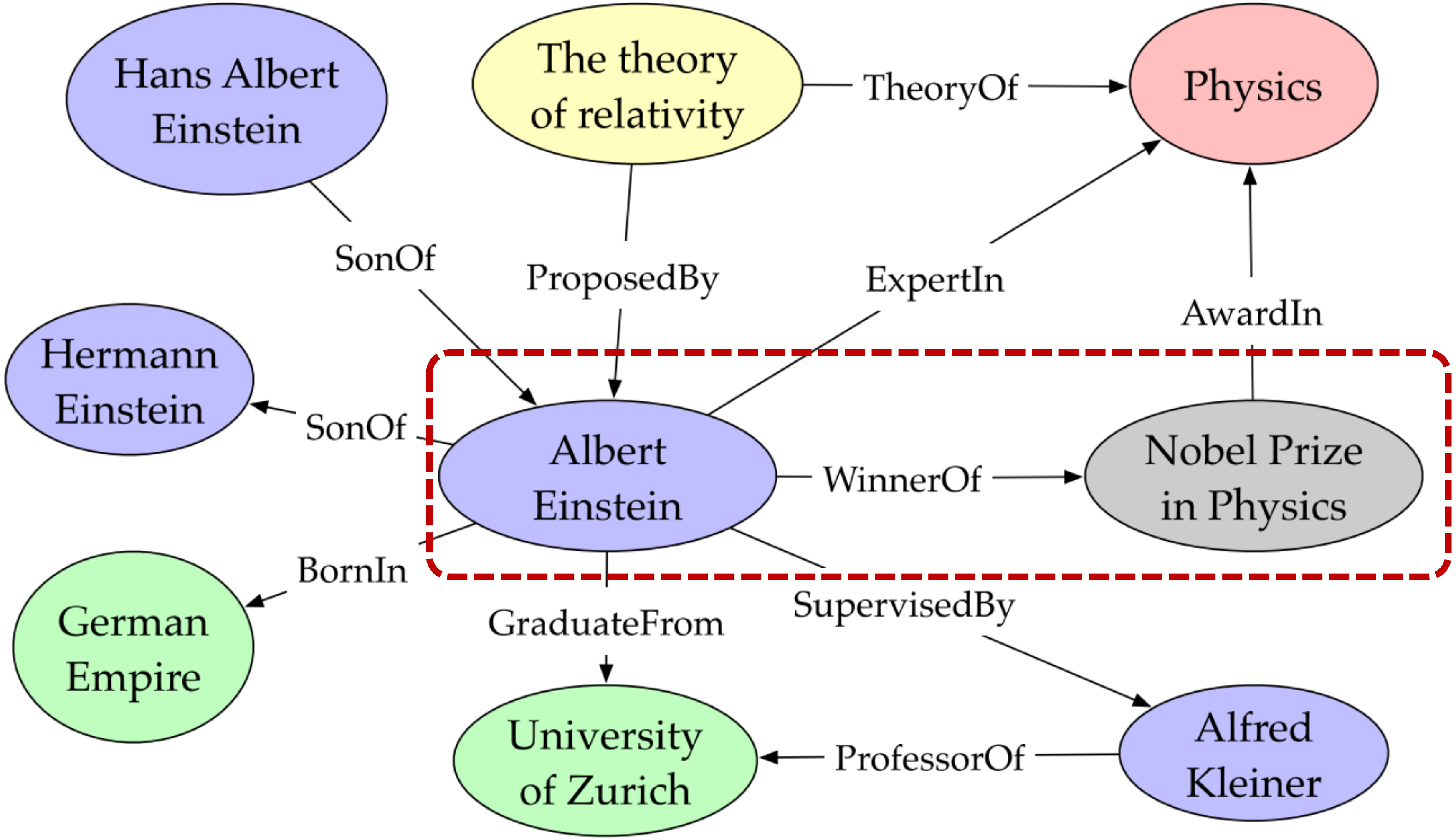
(h, r, t)

(Albert Einstein, **BornIn**, German Empire)
(Albert Einstein, **SonOf**, Hermann Einstein)
(Albert Einstein, **GraduateFrom**, University of Zurich)
(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
(Albert Einstein, **ExpertIn**, Physics)
(Nobel Prize in Physics, **AwardIn**, Physics)
(The theory of relativity, **TheoryOf**, Physics)
(Albert Einstein, **SupervisedBy**, Alfred Kleiner)
(Alfred Kleiner, **ProfessorOf**, University of Zurich)
(The theory of relativity, **ProposedBy**, Albert Einstein)
(Hans Albert Einstein, **SonOf**, Albert Einstein)

Source: Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems.

Entities and Relations in Knowledge Graph

(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)



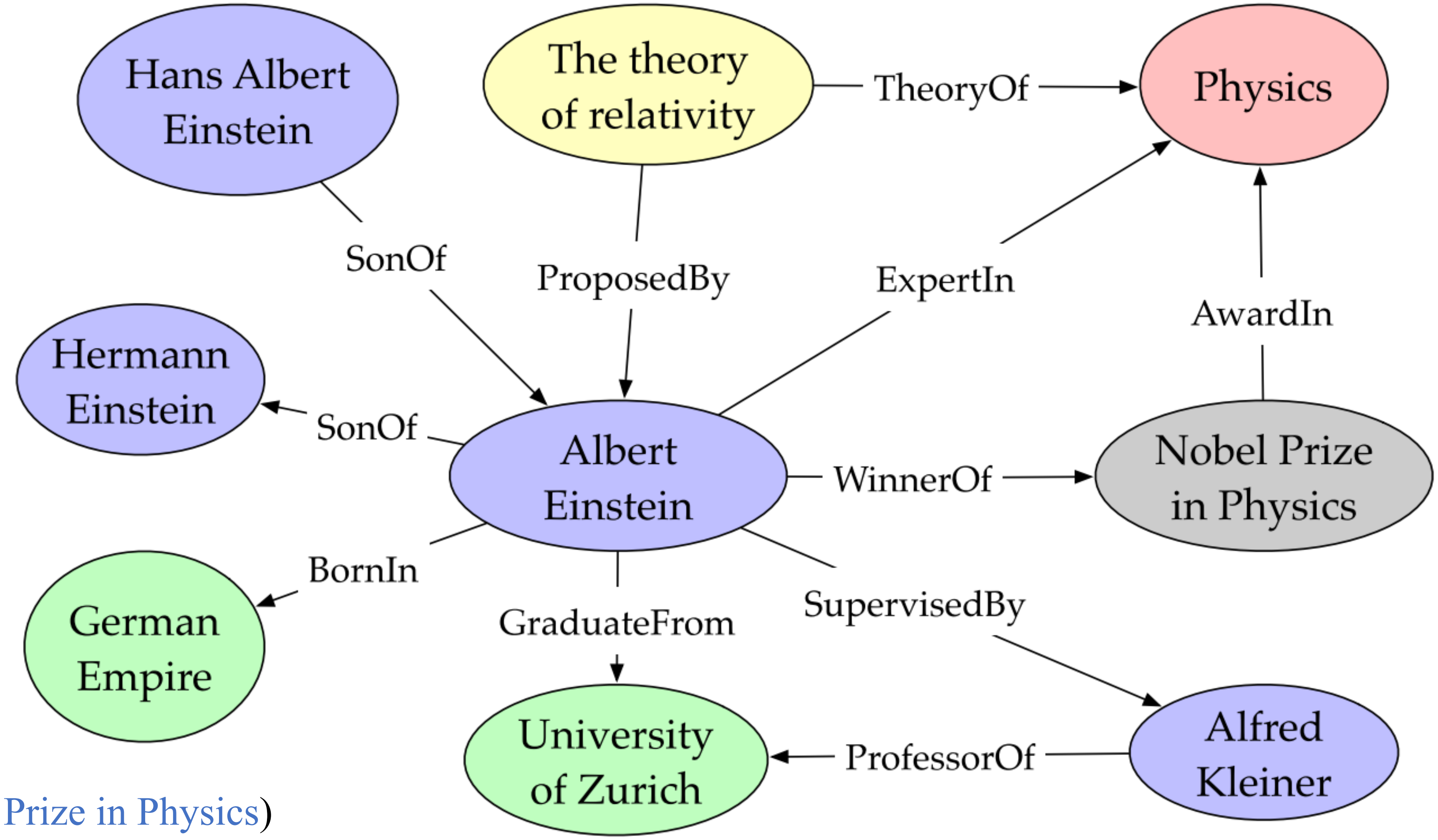
Source: Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems.

knowledge base and knowledge graph

Factual triples in knowledge base

(Albert Einstein, **BornIn**, German Empire)
(Albert Einstein, **SonOf**, Hermann Einstein)
(Albert Einstein, **GraduateFrom**, University of Zurich)
(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
(Albert Einstein, **ExpertIn**, Physics)
(Nobel Prize in Physics, **AwardIn**, Physics)
(The theory of relativity, **TheoryOf**, Physics)
(Albert Einstein, **SupervisedBy**, Alfred Kleiner)
(Alfred Kleiner, **ProfessorOf**, University of Zurich)
(The theory of relativity, **ProposedBy**, Albert Einstein)
(Hans Albert Einstein, **SonOf**, Albert Einstein)

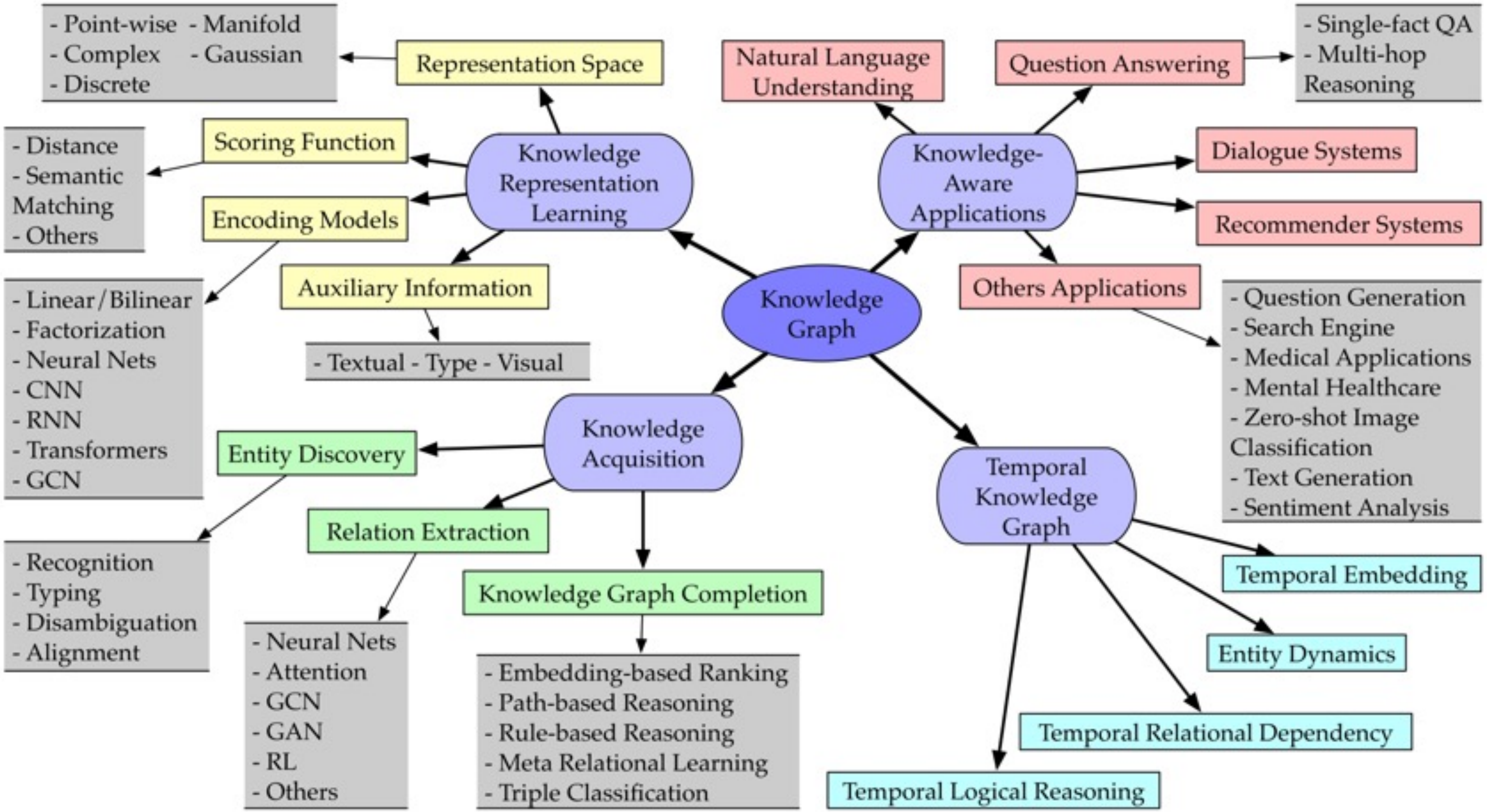
Entities and relations in knowledge graph



(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)

Source: Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems.

Categorization of Research on Knowledge Graphs



Source: Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems.

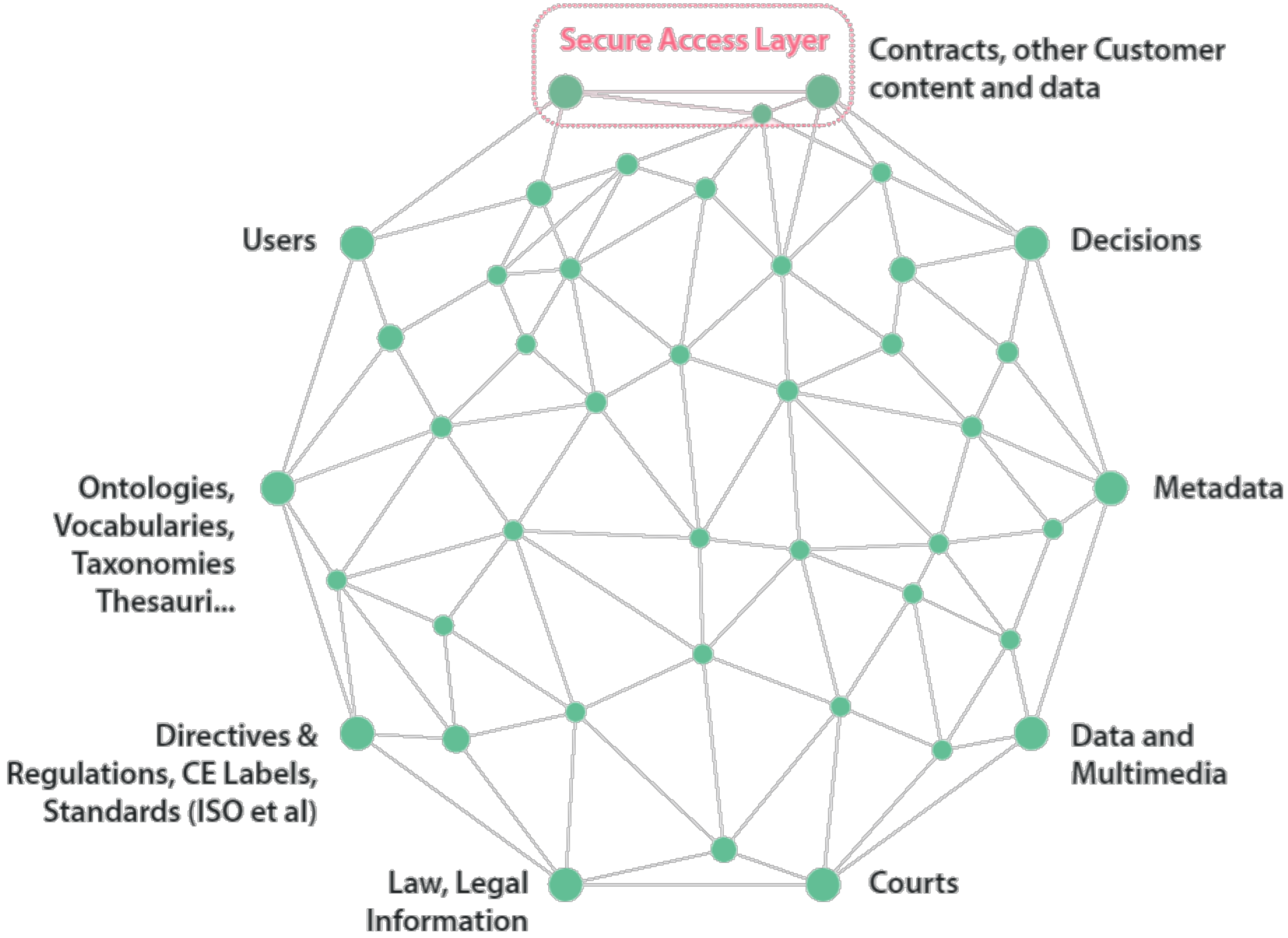
Knowledge Graph Completion (KGC) Datasets

Knowledge Graph Completion (KGC) Dataset	#Entity	#Relation	#Train	#Valid	#Test	Reference
WN18RR	40,943	11	86,835	3,034	3,134	Toutanova & Chen (2015); Zhang et al. (2020)
FB15k-237	14,541	237	272,115	17,535	20,466	Dettmers et al. (2018); Zhang et al. (2020)
YAGO3-10	123,182	37	1,079,040	5,000	5,000	Mahdisoltani et al. (2015); Zhang et al. (2020)

Domain-Specific Knowledge Graph

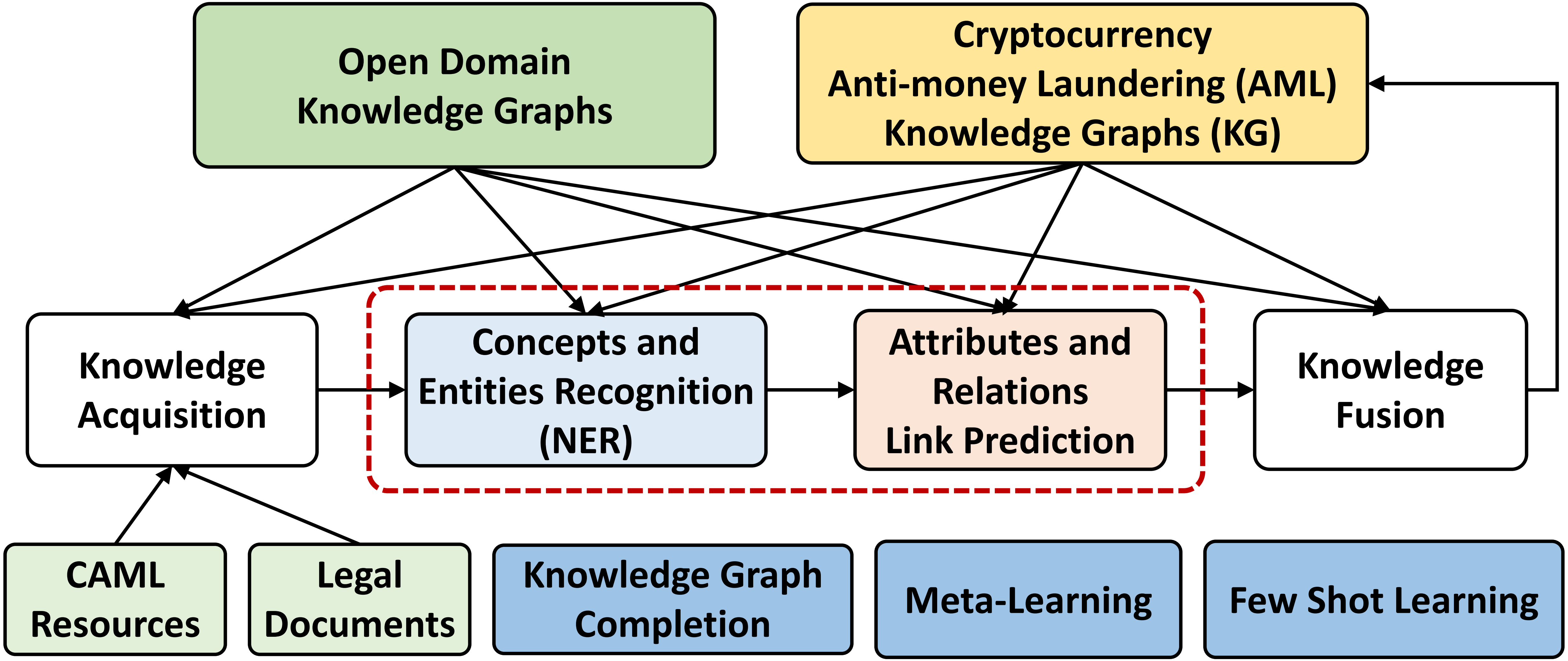
- **Domain-Specific Knowledge Graph**
  - **PubMed Knowledge Graph (PKG)**
    - **Extracting biological entities from 29 million PubMed abstracts**
  - **Lynx: Legal Knowledge Graph for Multilingual Compliance Services**
    - Legal Knowledge Graph (LKG) integrates and links heterogeneous compliance data sources including legislation, case law, standards and other private contracts.

Lynx: Legal Knowledge Graph for Multilingual Compliance Services



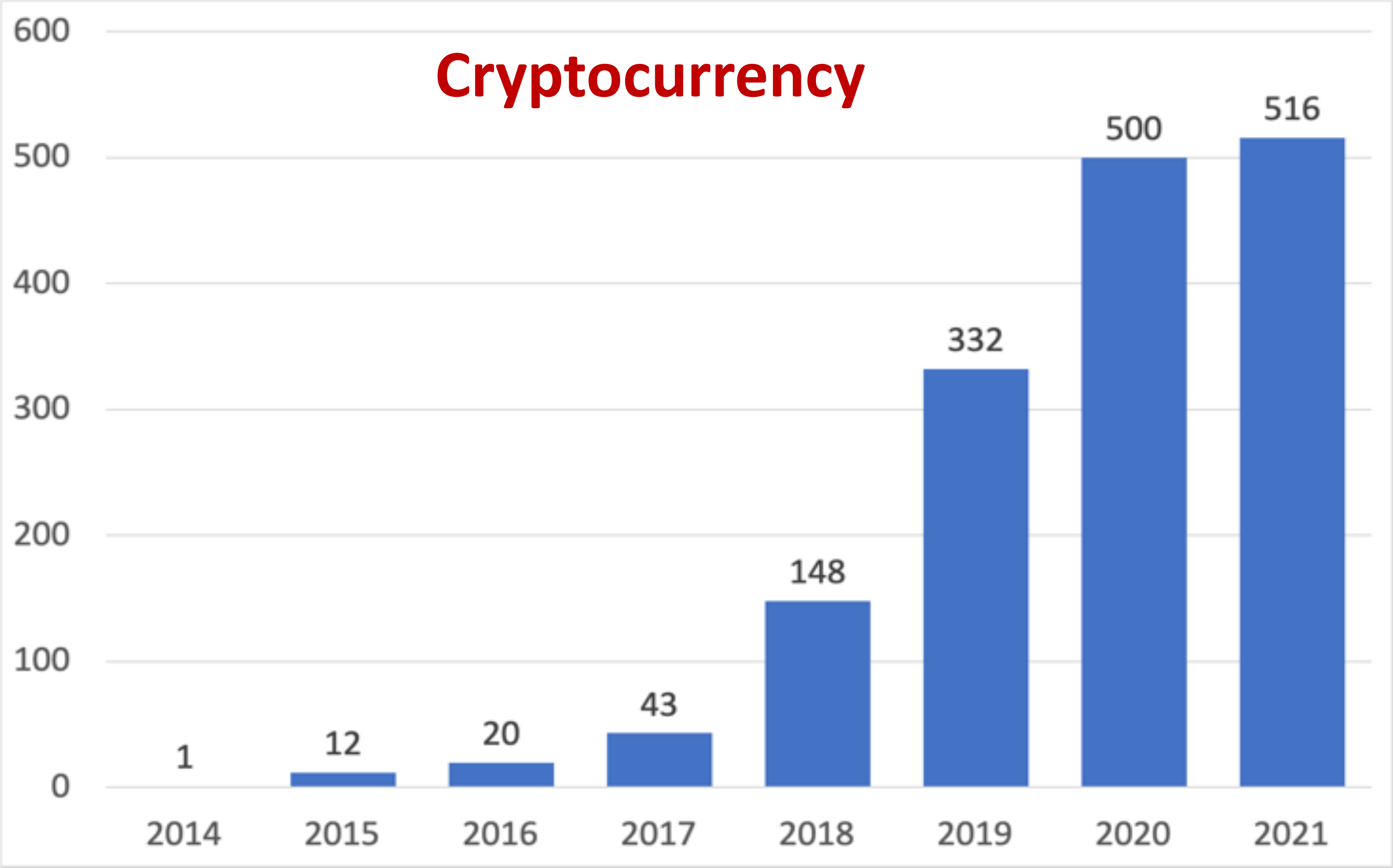
Source: Lynx Legal Knowledge Graph (LKG), https://lynx-project.eu/

System Architecture for Cryptocurrency Anti-money Laundering (AML) Knowledge Graphs



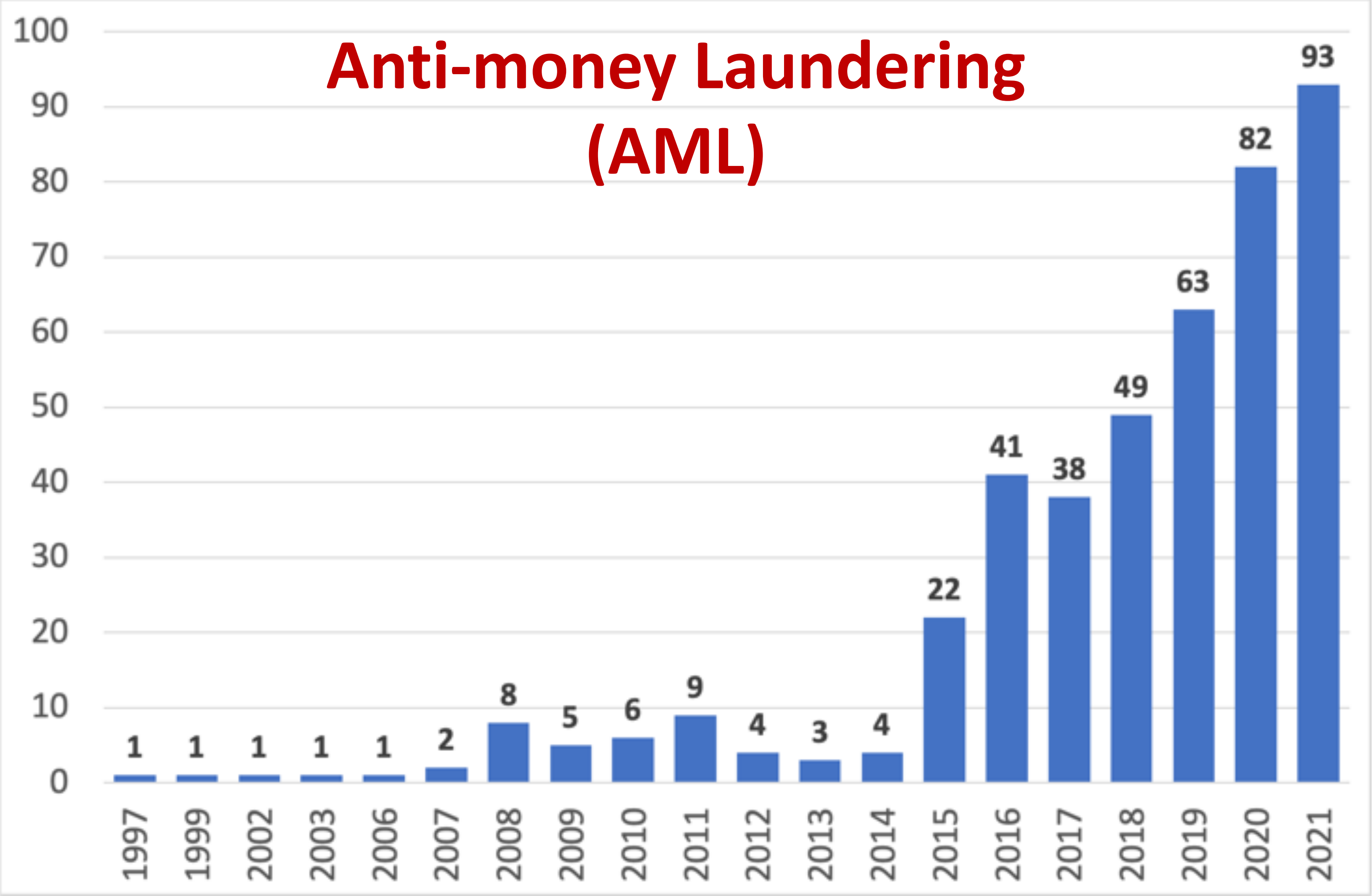
Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech", in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Research Trend of Cryptocurrency Research on Web of Science (2014-2021)



Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech", in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Research Trend of Anti-money Laundering Research on Web of Science (1997-2021)



Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech", in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Top keywords in Cryptocurrency

Rank	Keyword	Frequency	Percentage
1	bitcoin	945	6.91%
2	cryptocurrency	825	6.03%
3	blockchain	523	3.82%
4	cryptocurrencies	247	1.81%
5	volatility	222	1.62%
6	inefficiency	148	1.08%
7	gold	128	0.94%
8	hedge	88	0.64%
9	ethereum	85	0.62%
10	economics	75	0.55%
11	returns	73	0.53%
12	security	63	0.46%
13	technology	59	0.43%
14	internet	56	0.41%
15	market	56	0.41%
16	risk	55	0.40%
17	safe haven	54	0.39%
18	smart contracts	53	0.39%
19	garch	51	0.37%
20	model	51	0.37%

Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech",
in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Word Cloud for Cryptocurrency

Top keywords in Anti-money Laundering

Rank	Keyword	Frequency	Percentage
1	money laundering	173	7.15%
2	anti-money laundering	91	3.76%
3	compliance	31	1.28%
4	aml	26	1.07%
5	corruption	23	0.95%
6	crime	20	0.83%
7	fatf	20	0.83%
8	terrorism	17	0.70%
9	risk	15	0.62%
10	governance	14	0.58%
11	regulation	13	0.54%
12	anti-money laundering (aml)	12	0.50%
13	blockchain	12	0.50%
14	financial action task force	12	0.50%
15	lawyers	12	0.50%
16	terrorism financing	12	0.50%
17	bitcoin	11	0.45%
18	customer due diligence	11	0.45%
19	financial crime	11	0.45%
20	global governance	11	0.45%

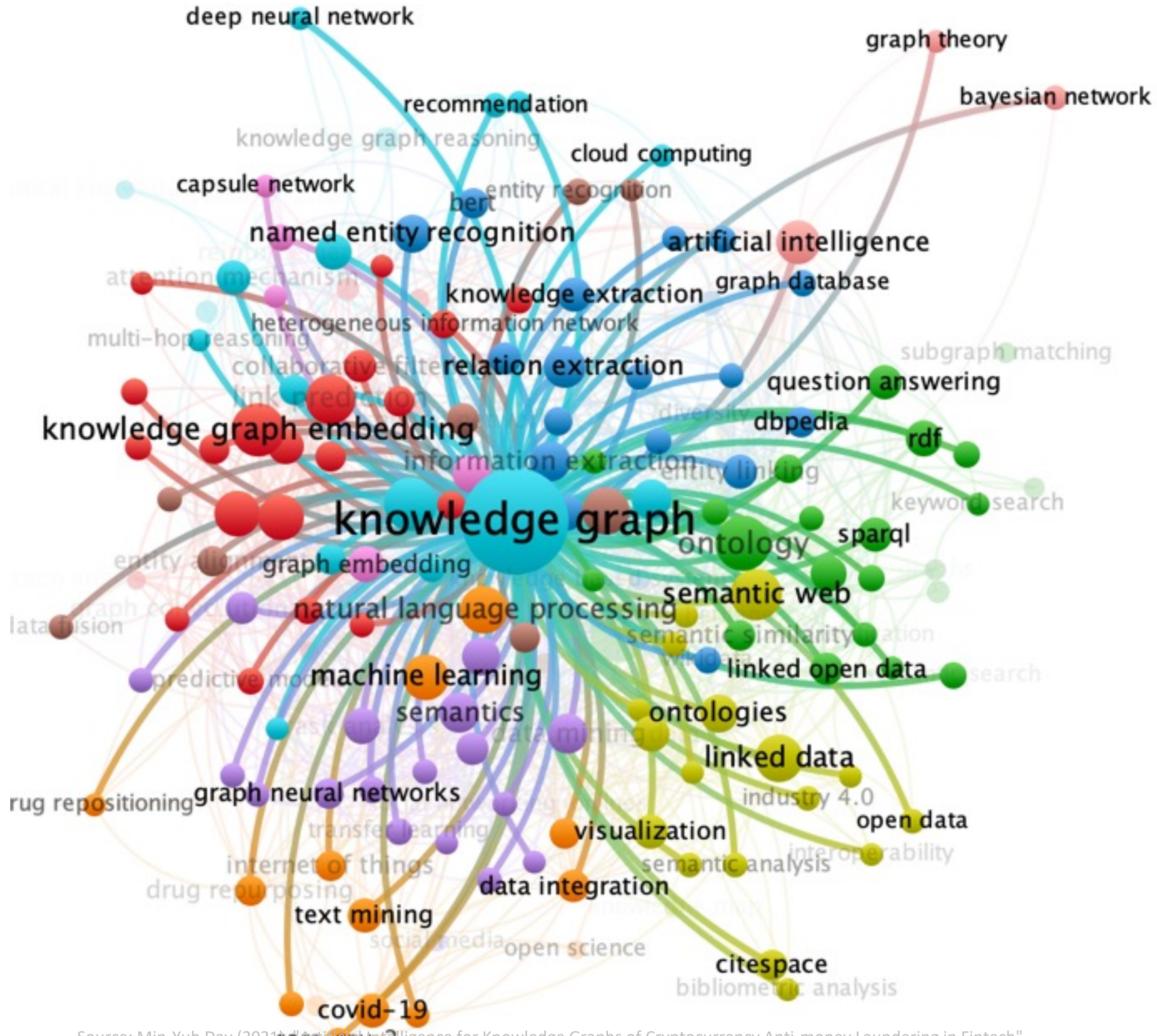
Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech", in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Word Cloud for Anti-money Laundering

Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech", in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Keyword Co-occurrence of Knowledge Graph

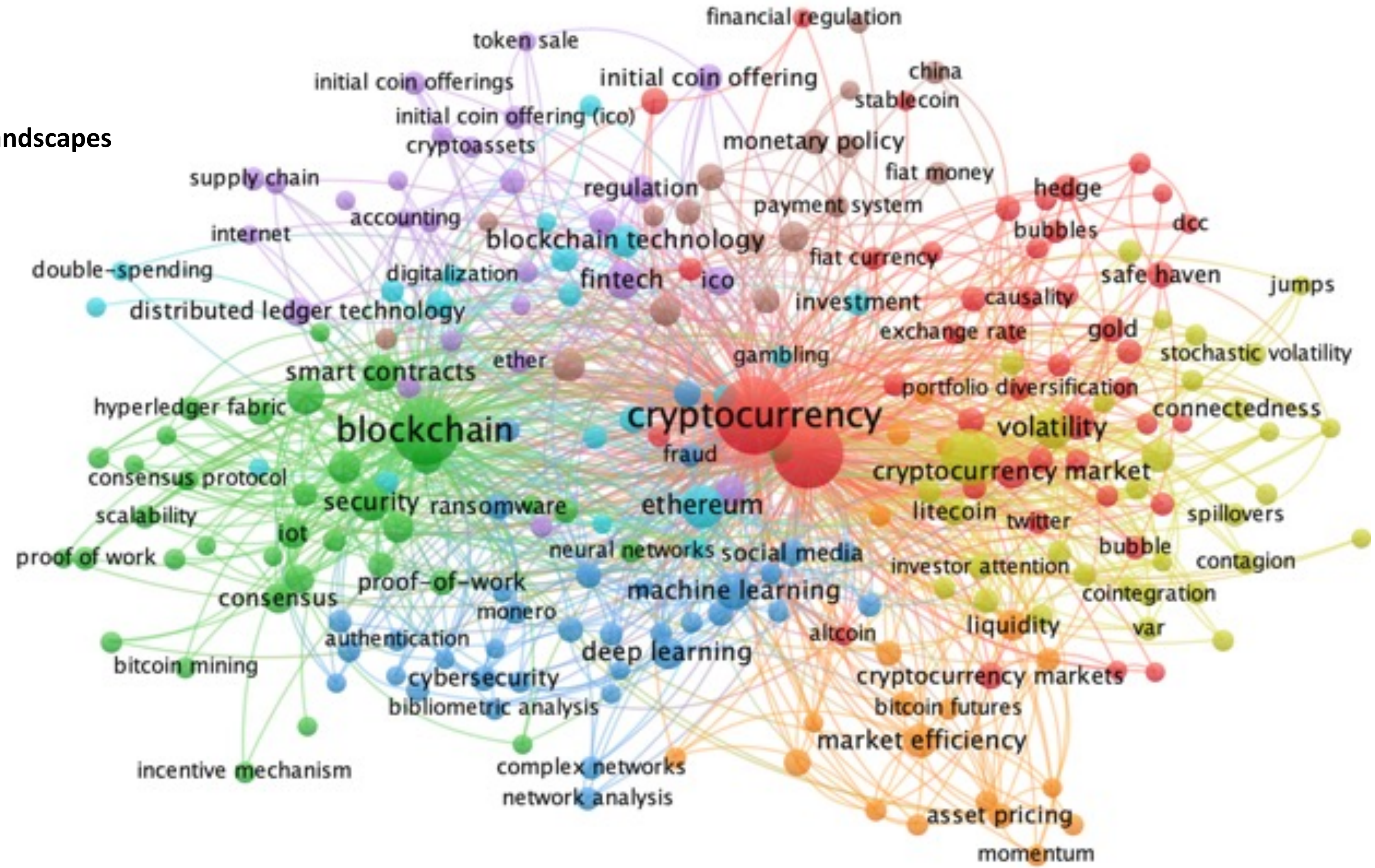
VOSviewer:
Visualizing scientific landscapes



Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech", in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Cryptocurrency

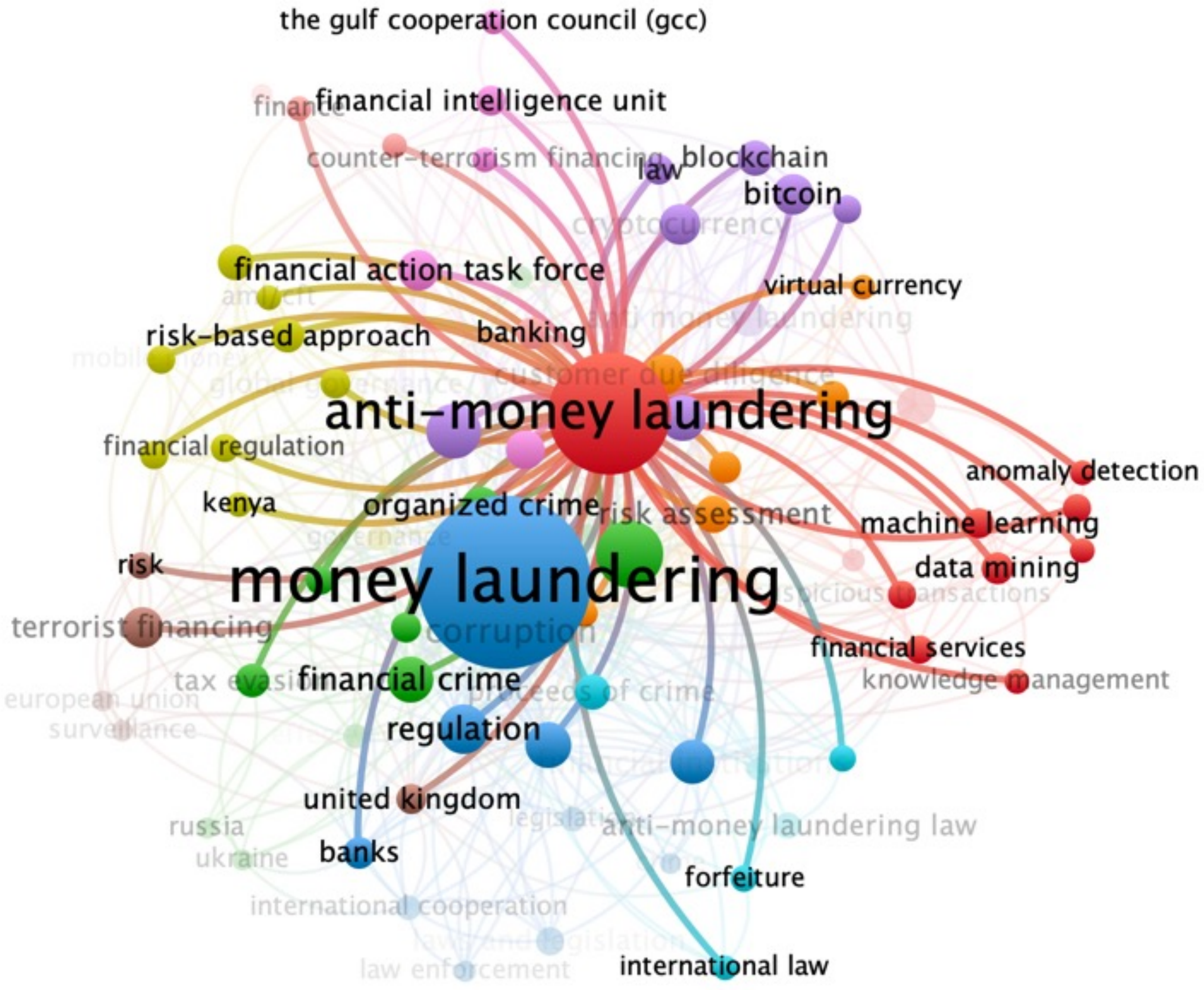
VOSviewer:
Visualizing scientific landscapes



Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech", in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Anti-money Laundering

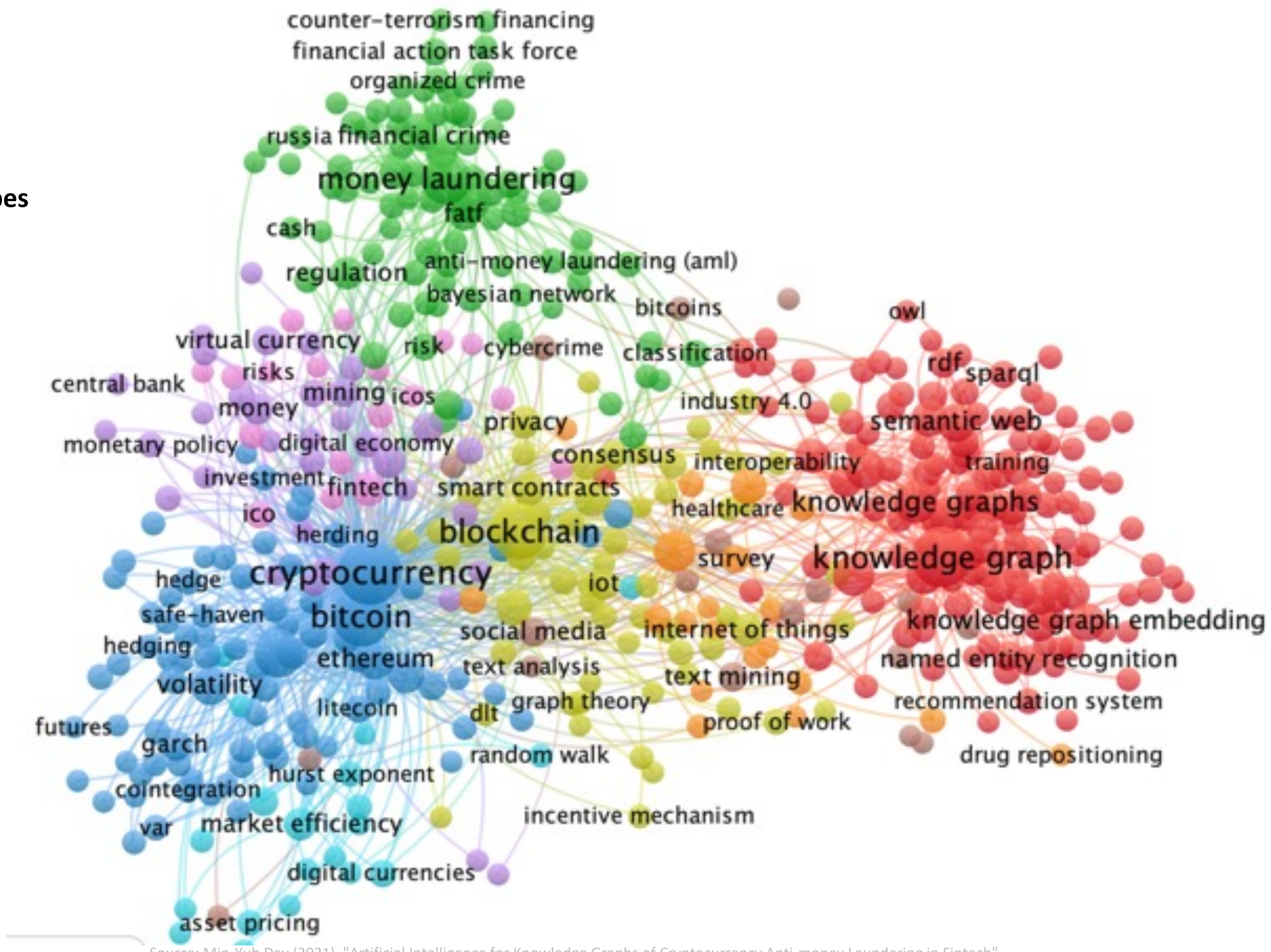
VOSviewer:
Visualizing scientific landscapes



Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech",
in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Knowledge Graphs, Cryptocurrency, Anti-money Laundering

VOSviewer:
Visualizing scientific landscapes



Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech",
in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Top 20 Journal Published Cryptocurrency Research on Web of Science (Total Journals: 1572)

Rank	Journal	Count	Percentage
1	Finance Research Letters	82	5.22%
2	IEEE Access	61	3.88%
3	Journal of Risk and Financial Management	40	2.54%
4	Physica a-Statistical Mechanics and Its Applications	39	2.48%
5	Economics Letters	33	2.10%
6	Research in International Business and Finance	29	1.84%
7	International Review of Financial Analysis	28	1.78%
8	Mathematics	21	1.34%
9	Applied Economics Letters	20	1.27%
10	Applied Economics	17	1.08%
11	Journal of Alternative Investments	16	1.02%
12	Frontiers in Blockchain	14	0.89%
13	Sustainability	13	0.83%
14	North American Journal of Economics and Finance	12	0.76%
15	Plos One	12	0.76%
16	Future Generation Computer Systems-the International Journal of Escience	11	0.70%
17	Journal of International Financial Markets Institutions & Money	11	0.70%
18	Journal of Money Laundering Control	11	0.70%
19	Technological Forecasting and Social Change	10	0.64%
20	Chaos Solitons & Fractals	9	0.57%

Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech",
in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Top 20 Journal Published Anti-money Laundering research on Web of Science (Total Journals: 160)

Rank	Journal	Count	Percentage
1	Journal of Money Laundering Control	197	45.39%
2	Crime Law and Social Change	18	4.15%
3	Trusts & Trustees	9	2.07%
4	Journal of Financial Regulation and Compliance	8	1.84%
5	European Journal on Criminal Policy and Research	6	1.38%
6	Financial and Credit Activity-Problems of Theory and Practice	6	1.38%
7	IEEE Access	4	0.92%
8	Journal of Banking Regulation	4	0.92%
9	International Journal of Disclosure and Governance	3	0.69%
10	International Review of Law and Economics	3	0.69%
11	Journal of Criminal Law	3	0.69%
12	Security Dialogue	3	0.69%
13	Arab Law Quarterly	2	0.46%
14	Asian Journal of Accounting and Governance	2	0.46%
15	Cuestiones Politicas	2	0.46%
16	Estudios De Economia Aplicada	2	0.46%
17	European Journal of Crime Criminal Law and Criminal Justice	2	0.46%
18	European Journal of Law and Economics	2	0.46%
19	Expert Systems with Applications	2	0.46%
20	Governance-an International Journal of Policy Administration and Institutions	2	0.46%

Source: Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech",
in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.

Summary

- AI for Text Analytics
  - Natural Language Processing with Transformers: Building Language Applications with Hugging Face
  - Practical Natural Language Processing
- FinTech: Financial Services Innovation
- Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech

Acknowledgments: Research Projects

- 計畫主持人，應用 AI 技術建構加密貨幣反洗錢知識圖譜：少樣本學習模型 (Applying AI technology to construct knowledge graphs of cryptocurrency anti-money laundering: a few-shot learning model)，
  - 科技部 (人文司 - 商事財經法)，110-2410-H-305-013-MY2，2021/08/01~2023/07/31 [核定經費 (新台幣)：1,022,000]
- 子計畫共同主持人，深化企業永續-由人工智慧、財務與策略觀點打造企業永續績效 (Deepen Corporate Sustainability: Enhance the Performance of Corporate Sustainability from AI, Financial, and Strategic Perspectives)：子計畫二：人工智慧企業永續評鑑與跨語言永續績效報告書生成式模型 (AI for Corporate Sustainability Assessment and Cross Language Corporate Sustainability Reports Generative Model)
  - 國立臺北大學，111-NTPU_ORDA-F-001，2022/01/01~2022/12/31 [經費總額 (新台幣)： 3,228,500]
- 子計畫主持人，人工智慧方法分析企業科技創新導入- 專利文字分析與影像分析應用 (Artificial intelligence methods applied for analyzing the introduction of technological innovation: Patent text analysis and image analysis)：子計畫三：應用人工智慧於專利文本分析金融科技知識圖譜 (Artificial Intelligence for FinTech Knowledge Graph from Patent Textual Analytics)
  - 國立臺北大學，111-NTPU_ORDA-F-003，2022/01/01~2022/12/31 [經費總額 (新台幣)： 1,291,950]
- 子計畫共同主持人，企業永續動機、價值攸關性與人工智慧於企業永續績效評比之應用 (Corporate Sustainability: Motivations, Value Relevance, and the Application of AI in the Assessment)：子計畫二：人工智慧 AI 於企業永續評比之應用 (An application of artificial intelligence (AI) in the corporate sustainability assessment)
  - 國立臺北大學，110-NTPU_ORDA-F-001，2021/01/01~2021/12/31 [經費總額 (新台幣)： 3,240,000]

Q & A

Artificial Intelligence for Text Analytics in FinTech (金融科技人工智慧文本分析)

Host: 林川傑 助理教授 (Prof. Chuan-Jie Lin)
Department of Computer Science and Engineering, National Taiwan Ocean University (NTOU)
Time: 13:00-15:00, May 19, 2022 (Thursday)
Place: Microsoft Teams, NTOU

戴敏育 副教授

Min-Yuh Day, Ph.D, Associate Professor

國立臺北大學 資訊管理研究所

Institute of Information Management, National Taipei University

https://web.ntpu.edu.tw/~myday

2022-05-19

References

- Lewis Tunstall, Leandro von Werra, and Thomas Wolf (2022), Natural Language Processing with Transformers: Building Language Applications with Hugging Face, O'Reilly Media.
- Denis Rothman (2021), Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more, Packt Publishing.
- Savaş Yıldırım and Meysam Asgari-Chenaghlu (2021), Mastering Transformers: Build state-of-the-art models from scratch with advanced natural language processing techniques, Packt Publishing.
- Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta (2020), Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems, O'Reilly Media.
- Campbell R. Harvey, Ashwin Ramachandran, Joey Santoro, Fred Ehrsam (2021), DeFi and the Future of Finance, Wiley
- Aurélien Géron (2019), Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd Edition, O'Reilly Media.
- Paolo Sironi (2016), FinTech Innovation: From Robo-Advisors to Goal Based Investing and Gamification, Wiley.
- Matt Fortnow and QuHarrison Terry (2021), The NFT Handbook - How to Create, Sell and Buy Non-Fungible Tokens, Wiley
- Parma Bains, Mohamed Diaby, Dimitris Drakopoulos, Julia Faltermeier, Federico Grinberg, Evan Papageorgiou, Dmitri Petrov, Patrick Schneider, and Nobu Sugimoto (2021), The Crypto Ecosystem and Financial Stability Challenges, International Monetary Fund, October 2021
- Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), "Business Intelligence, Analytics, and Data Science: A Managerial Perspective", 4th Edition, Pearson
- Frederic S. Mishkin (2015), "The Economics of Money, Banking and Financial Markets", 11th Edition, Pearson
- Xiao-lin Zheng, Meng-ying Zhu, Qi-bing Li, Chao-chao Chen, and Yan-chao Tan (2019), "Finbrain: When finance meets AI 2.0." Frontiers of Information Technology & Electronic Engineering 20, no. 7, pp. 914-924
- Liye Ma and Baohong Sun (2020), "Machine learning and AI in marketing – Connecting computing power to human insights." International Journal of Research in Marketing, 37, no. 3, 481-504.
- Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer (2020). "Deep learning for financial applications: A survey." Applied Soft Computing (2020): 106384.
- Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu (2020), "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019." Applied Soft Computing 90 (2020): 106181.
- Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). "A survey on knowledge graphs: Representation, acquisition, and applications." IEEE Transactions on Neural Networks and Learning Systems.
- Min-Yuh Day, Yirung Cheng, Paoyu Huang, and Yensen Ni (2022), "The profitability of trading US stocks in Quarter 4 - evidence from trading signals emitted by SOI and RSI", Applied Economics Letters, 17 February 2022, pp. 1-6.
- Kuei-Hui Cheng, Min-Yuh Day, and Hsiao-Lun Lin (2021), "The Impact of Revenue Key Audit Matters on the Difference between Self-declared and Audited Revenue", Taiwan Accounting Review, Volume 17, Number 2, December 2021, pp. 309-356.
- Yulu Liao, Min-Yuh Day, Yirung Cheng, Paoyu Huang, and Yensen Ni (2021), "The profitability of Technical Trading for Hotel Stocks Under COVID-19 Pandemic", Journal of Computers, Volume 32, Number 5, October 2021, pp. 44-54.
- Yulu Liao, Min-Yuh Day, Yirung Cheng, Paoyu Huang, and Yensen Ni (2021), "Does CBOE volatility index jumped or located at a higher level matter for evaluating DJ 30, NASDAQ, and S&P500 index subsequent performance", Journal of Computers, Volume 32, Number 4, August 2021, pp. 57-66.
- Min-Yuh Day, Pao-Yu Huang, Yirung Cheng, Yin-Tzu Lin, and Yensen Ni (2021), "Profitable day trading Bitcoin futures following continuous bullish (bearish) candlesticks", Applied Economics Letters, 16 March 2021, pp. 1-8.
- Yensen Ni, Min-Yuh Day, and Paoyu Huang (2020), "Trading Stocks Following Sharp Movements in the USDX, GBP/USD, and USD/CNY", Financial Innovation, Volume 6, 35, 2020, pp. 1-17.
- Yensen Ni, Min-Yuh Day, Paoyu Huang, and Shang-Ru Yua (2020), "The profitability of Bollinger Bands: Evidence from the constituent stocks of Taiwan 50", Physica A: Statistical Mechanics and its Applications, Volume 551, 1 August 2020, 124144, pp. 1-14.
- Yensen Ni, Paoyu Huang, Yaochia Ku, Yiching Liao, and Min-Yuh Day (2020), "Investing Strategies as Stochastic Oscillator Indicators Staying in Overreaction Zones for Consecutive Days with Big Data Concerns", Journal of Computers, Volume 31, Number 1, February 2020, pp. 1-17.
- Yensen Ni, Manhwa Wu, Min-Yuh Day, and Paoyu Huang (2020), "Do sharp movements in oil prices matter for stock markets?", Physica A: Statistical Mechanics and its Applications, Volume 539, 1 February 2020, pp. 1-11.
- Min-Yuh Day, Paoyu Huang, and Yensen Ni (2019), "Trading as sharp movements in oil prices and technical trading signals emitted with big data concerns", Physica A: Statistical Mechanics and its Applications, Volume 525, 1 July 2019, pp. 349-372.
- Yensen Ni, Min-Yuh Day, and Paoyu Huang (2019), "Does Data Frequency Matter for Trading Signals Emitted by Various Technical Trading Rules? ", Pacific Business Review International, Volume 11, Issue 10, April 2019, pp. 7-17.
- Min-Yuh Day, Manhwa Wu, Paoyu Huang, and Yensen Ni (2018), "Investing Strategies as a Sharp Movement in Exchange Rates Occurred– Evidence for the Constituent Stocks of SSE 50 and TW 50", The Journal of Investing, Volume 27, Issue 4, Winter 2018, pp. 58-68.
- Min-Yuh Day, Paoyu Huang, Yensen Ni, and Yuhsin Chen (2018), "Do Implicit Phenomena Matter? Evidence from China Stock Index Futures", The Journal of Alternative Investments, Volume 21, Issue 1, Summer 2018, pp. 79-91.
- Yensen Ni, Yirung Cheng, Paoyu Huang, and Min-Yuh Day (2018), "Trading strategies in terms of continuous rising (falling) prices or continuous bullish (bearish) candlesticks emitted", Physica A: Statistical Mechanics and its Applications, Volume 501, 1 July 2018, pp. 188-204.
- Min-Yuh Day, Paoyu Huang, Yensen Ni, and Yuhsin Chen (2018), "Do Intraday Large Price Changes Matter for Trading Index Futures? Evidence from China Futures Markets", Journal of Financial Studies, Volume 26, Number 2, June 2018, pp. 139-174.
- Min-Yuh Day (2021), "Artificial Intelligence for Knowledge Graphs of Cryptocurrency Anti-money Laundering in Fintech", in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2021), Virtual Event, Netherlands, November 8-11, 2021.
- Min-Yuh Day and Jian-Ting Lin (2019), "Artificial Intelligence for ETF Market Prediction and Portfolio Optimization", in Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2019), Vancouver, Canada, August 27-30, 2019.
- Min-Yuh Day,Tun-Kung Cheng and Jheng-Gang Li (2018), "AI Robo-Advisor with Big Data Analytics for Financial Services", in Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2018), Barcelona, Spain, August 28-31, 2018.
- Min-Yuh Day, Jian-Ting Lin and Yuan-Chih Chen (2018), "Artificial Intelligence for Conversational Robo-Advisor", in Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2018), Barcelona, Spain, August 28-31, 2018.
- Min-Yuh Day and Chia-Chou Lee (2016), "Deep Learning for Financial Sentiment Analysis on Finance News Providers", in Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2016), San Francisco, California, USA, Aug 18-21, 2016.